

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:50:28 ; Search time 97.9924 Seconds

(without alignments)
2185.588 Million cell updates/sec

Title: US-09-903-199-2

Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IYDVWHPDLTPQRRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4022	100.0	758	4 AAB83919	Aab83919 A human a
2	4022	100.0	758	6 ABG72365	Abg72365 Human asp
3	4022	100.0	758	6 ADA00639	Ada00639 Human asp
4	1353.5	33.7	265	6 ABU92053	Abu92053 Human pro
5	1334	33.2	255	2 AAY33642	Aay33642 Human lab
6	1320	32.8	255	5 AAU85544	Aau85544 Clone #48
7	1320	32.8	255	6 ABU69516	Abu69516 Human lun
8	1320	32.8	255	6 ABU66419	Abu66419 Lung canc
9	933	23.2	422	4 ABB61986	Abb61986 Drosophil
10	274.5	6.8	369	4 AAB73682	Aab73682 Human oxi
11	271.5	6.8	236	3 AAB43327	Aab43327 Human ORF
12	262	6.5	264	6 ABM68322	Abm68322 Photorhab
13	241	6.0	104	4 AAU29679	Aau29679 Novel hum
14	218.5	5.4	299	6 ABU21099	Abu21099 Protein e
15	217.5	5.4	324	4 AAU28081	Aau28081 Novel hum
16	215	5.3	47	4 AAU31979	Aau31979 Novel hum
17	204.5	5.1	311	4 ABB60327	Abb60327 Drosophil
18	203.5	5.1	562	2 AAR70491	Aar70491 Leucocyto
19	199	4.9	109	4 AAB73672	Aab73672 Human oxi
20	199	4.9	783	2 AAR05804	Aar05804 C-termina
21	197.5	4.9	1018	2 AAR98747	Aar98747 P. vivax
22	197.5	4.9	1018	2 AAW97039	Aaw97039 A secreta
23	197.5	4.9	1018	4 AAG66528	Aag66528 Plasmodiu
24	197.5	4.9	1132	2 AAR97866	Aar97866 Chicken I
25	196.5	4.9	355	6 ABU23196	Abu23196 Protein e

26	191.5	4.8	1881	3 AAY44506	Aay44506 Streptoco
27	191.5	4.8	1881	6 ABU01047	Abu01047 S. pneumo
28	190.5	4.7	778	3 AAG46504	Aag46504 Arabidops
29	190.5	4.7	788	4 ABB68264	Abb68264 Drosophil
30	190.5	4.7	1162	3 AAY96255	Aay96255 Kaposi's
31	190.5	4.7	1162	3 AAY58500	Aay58500 HHV8 ORF
32	190.5	4.7	1162	4 AAB62331	Aab62331 Amino aci
33	190.5	4.7	1162	5 ABB05621	Abb05621 Kaposi's
34	190	4.7	360	2 AAW03627	Aaw03627 Human fol
35	189.5	4.7	304	6 ABU17264	Abu17264 Protein e
36	189.5	4.7	312	6 ADA36728	Ada36728 Acinetoba
37	189.5	4.7	1616	6 ABU35669	Abu35669 Protein e
38	189.5	4.7	1616	7 ABO23515	Abo23515 Mycoplasma
39	189	4.7	3263	4 ABB67210	Abb67210 Drosophil
40	187.5	4.7	1192	4 ABB59642	Abb59642 Drosophil
41	186.5	4.6	412	2 AAW03626	Aaw03626 Human thy
42	186	4.6	49	6 ADA56732	Ada56732 Human sec
43	186	4.6	49	6 ADA40577	Ada40577 Human sec
44	186	4.6	49	6 ABR47638	Ab47638 Human sec
45	186	4.6	50	2 AAY01148	Aay01148 Secreted

ALIGNMENTS

RESULT 1	
AAB83919	
ID AAB83919 standard; protein; 758 AA.	
XX	
AC AAB83919;	
XX	
DT 23-JUL-2001 (first entry)	
XX	
DE A human aspartyl (asparaginy1) beta-hydroxylase (HAAH).	
XX	
KW Epidermal growth factor-like domain; EGF-like domain; cancer;	
KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.	
XX	
OS Homo sapiens.	
XX	
PN WO200135102-A2.	
XX	
PD 17-MAY-2001.	
XX	
PF 08-NOV-2000; 2000WO-US030738.	
XX	
PR 08-NOV-1999; 99US-00436184.	
XX	
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.	
XX	
PI Wands JR, De la Monte SM, Ince N, Carlson RI;	
XX	
DR WPI; 2001-329171/34.	
DR N-PSDB; AAF89811.	
XX	
PT Diagnosing malignant neoplasm in a mammal, involves contacting mammalian	
PT sample with antibody that binds to human aspartyl beta-hydroxylase	
PT polypeptide to form antigen-antibody complex and detecting the complex.	
XX	
PS Disclosure; Page 5; 76pp; English.	
XX	
CC The present sequence represents a human aspartyl (asparaginy1) beta-	
CC hydroxylase (HAAH) enzyme. Epidermal growth factor (EGF)-like domains of	
CC polypeptides are hydroxylated by HAAH enzymes. HAAH is used in the method	
CC of the invention. The specification describes a method for diagnosing a	
CC malignant neoplasm in a mammal. The method comprises contacting a body	
CC fluid with an antibody which binds to HAAH polypeptide under complex	
CC forming conditions, and detecting the antigen-antibody complex. The	
CC method is useful for diagnosing and prognosing a malignant neoplasm in a	
CC bodily fluid e.g. central nervous system (CNS)-derived body fluid, blood,	
CC serum, urine, saliva, sputum, lung effusion, and ascites fluid of mammal,	
CC where the neoplasm is derived from endodermal tissue and is selected from	
CC colon cancer, breast cancer, pancreatic cancer, liver cancer, cancer of	

CC bile ducts, and cancer of the CNS. HAAH antibodies, linked to a cytotoxic
CC agent, are useful for killing tumour cells
XX
XX
SQ Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 4; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.1e-311;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGSSSSGSGSSTAGSSSPGARRETKHGHNKGRKGLSGTSFTTWNV 60
Db 1 MAQRKNAKSSGSSSSGSGSSTAGSSSPGARRETKHGHNKGRKGLSGTSFTTWNV 60
QY 61 IALLGWTSVAVVWFPLVDYEEVLGKLGIVDADGDGDFDVAADAKVLLGLKERSTSEPAVP 120
Db 61 IALLGWTSVAVVWFPLVDYEEVLGKLGIVDADGDGDFDVAADAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEHPTEPEEQVPEAEPPQNIIDEAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGEPO 180
Db 121 PEEAEHPTEPEEQVPEAEPPQNIIDEAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGEPO 180
QY 181 QEDDEFMATDVDRFETLEPEVSHETEHSYHVEETVSQDCNQDMEMMSQENPDSE 240
Db 181 QEDDEFMATDVDRFETLEPEVSHETEHSYHVEETVSQDCNQDMEMMSQENPDSE 240
QY 241 PVEDERLHNDTDVITYOYBEQAVPEPLENEGIEITEVTAPPEPNVEDSQVIVEEVS 300
Db 241 PVEDERLHNDTDVITYOYBEQAVPEPLENEGIEITEVTAPPEPNVEDSQVIVEEVS 300
QY 301 FVEEQEQEVPETNRKTDDPEQAKAVKKKKPKLINKFDKTIKAEIDAEKLRKRGKIEEA 360
Db 301 FVEEQEQEVPETNRKTDDPEQAKAVKKKKPKLINKFDKTIKAEIDAEKLRKRGKIEEA 360
QY 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETVQEVASLPDVPADLLK 420
Db 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETVQEVASLPDVPADLLK 420
QY 421 LSLKRSDRQOFLGHNRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
Db 421 LSLKRSDRQOFLGHNRGSLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
QY 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPSTDGGRFYFHLGDAMQRYEN 540
Db 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPSTDGGRFYFHLGDAMQRYEN 540
QY 541 KEAYKMYELGHRGHPASVQWQSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKMYELGHRGHPASVQWQSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRRENENACKGAPKCTTLLKFPETTG 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRRENENACKGAPKCTTLLKFPETTG 660
QY 661 RRGQIKYSIMHPGTHVWPHPTGPTNCLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
Db 661 RRGQIKYSIMHPGTHVWPHPTGPTNCLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 758
Db 721 DSFEHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 758

RESULT 2
ID ABG72365
ID ABG72365 standard; protein; 758 AA.
AC ABG72365;
XX 06-FEB-2003 (first entry)
DT
XX Human aspartyl (asparaginyl) beta-hydroxylase, HAAH.
DE Human; enzyme; aspartyl (asparaginyl) beta-hydroxylase; HAAH; cytostatic;
XX

KW immunostimulant; antibody; neoplasm; tumour; FB50; 86A; 5C7; 19B;
KW brain tumour; glioma; glioblastoma; astrocytoma; haemangioma;
KW pancreatic carcinoma; colon cancer; breast cancer; pancreatic cancer;
KW liver cancer; cancer of the bile ducts; primary malignant CNS neoplasm;
KW metastatic CNS neoplasm.
XX
XX Homo sapiens.
OS

PH Key Location/Qualifiers
FT Binding-site 286..291
FT /label= Antibody binding_site
FT /note= "This region is specifically claimed in claim 36"
FT Binding-site 573..579
FT /label= Antibody binding_site
FT /note= "This region is specifically claimed in claim 37"
FT Domain 600..700
FT /label= Catalytic domain 1
FT /note= "This region is specifically claimed in claim 7"
FT Binding-site 613..620
FT /label= Antibody binding_site
FT /note= "This region is specifically claimed in claim 38"
FT Domain 650..700
FT /label= Catalytic domain 2
FT /note= "This region is specifically claimed in claim 24"
FT Binding-site 660..758
FT /note= "Alpha ketoglutarate binding site. This region is specifically claimed in claim 32"
FT Domain 679..697
FT /note= "Epidermal growth factor-like domain; This region is specifically claimed in claim 33"

US2002110559-A1.

15-AUG-2002.

17-MAY-2001; 2001US-00859604.

08-NOV-1999; 99US-00436184.

(WAND/) WANDS J R.
(DMON/) DE LA MONTE S M.

(DEUT/) DEUTCH A H.
(GHAN/) GHANBARI H A.

Wands JR, De La Monte SM, Deutch AH, Ghanbari HA;

WPI; 2003-066676/06.

N-PSDB; ABS57868.

PT Diagnosing neoplasm and inhibiting tumor growth, by contacting tissue of
PT mammal with detectably-labeled antibody which binds to human aspartyl
PT (asparaginyl) beta-hydroxylase.
XX

Claim 14; Page 3; 34pp; English.

XX The invention relates to diagnosing a neoplasm and inhibiting tumour
XX growth in a mammal, using an antibody that binds to human aspartyl
CC (asparaginyl) beta-hydroxylase (HAAH). Diagnosing a neoplasm comprises
CC contacting a tissue with a detectably-labelled antibody where an increase
CC in level of antibody binding at tissue site compared to the level of
CC binding to normal non-neoplastic tissue indicates the presence of a
CC neoplasm at the tissue site. Inhibiting tumour growth in mammal involves
CC administering the antibody conjugated to a cytotoxic agent to a mammal.
CC Also included are a method of conferring an immune response to a tumour
CC cell in a mammal, by administering the antibody, a method of inducing an
CC HAAH-specific immune response in a mammal, by administering to the mammal
CC an HAAH polypeptide (or a polynucleotide composition encoding the
CC polypeptide, or its degenerate variant), a fragment of HAAH comprising an
CC extracellular domain and lacking a cytoplasmic domain of HAAH, an
CC antibody or its fragment which binds to HAAH (where the antibody is FB50,
CC 86A, 5C7 or 19B), a hybridoma cell line chosen from hybridoma FB50,
CC HA386A, HA15CA and HA219B, and a fragment of HAAH which lacks enzymatic
CC activity or alpha-ketoglutarate binding domain and epidermal growth

CC factor (EGF)-like domain. The methods are useful for diagnosing neoplasm
CC in a mammal, inhibiting tumour growth in a mammal, conferring an immune
CC response to a brain tumour cell (e.g. glioma, glioblastoma, astrocytoma
CC or haemangioma) in a mammal, for conferring immune response to a
CC pancreatic carcinoma cell and for inducing a HAAH-specific immune
CC response in a mammal. The method is useful for diagnosing malignant
CC neoplasms derived from endodermal tissue, e.g. colon cancer, breast
CC cancer, pancreatic cancer, liver cancer and cancer of the bile ducts. The
CC method is also useful for diagnosing neoplasms of central nervous system
CC (CNS) e.g. primary malignant CNS neoplasms of both neuronal and glial
CC cell origin and metastatic CNS neoplasms, and for diagnosing brain
CC tumours e.g. glioma, glioblastoma, astrocytoma or haemangioma. The
CC present sequence represents human HAAH
XX
SQ Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.1e-311;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGNSSSSGSGSGSSTAGSSSPGARRETKHGKNGRKGLSGTSFFTFMWV 60
DB 1 MAQRKNAKSSGNSSSSGSGSGSSTAGSSSPGARRETKHGKNGRKGLSGTSFFTFMWV 60
QY 61 IALLGWTSAVAVVWFDLVDYEEVGLKIGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
DB 61 IALLGWTSAVAVVWFDLVDYEEVGLKIGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEPQNIIEDEAKEQIQSLHEWVHAHEVGEDLQOEDGPTGEPQ 180
DB 121 PEEAEPHTEPEEQVPVEAEPQNIIEDEAKEQIQSLHEWVHAHEVGEDLQOEDGPTGEPQ 180
QY 181 QEDDEFLMATDVDDRFETLEPEVSHETESHYHVEETVSQDCNQDMEEMMSEQENPDSSSE 240
DB 181 QEDDEFLMATDVDDRFETLEPEVSHETESHYHVEETVSQDCNQDMEEMMSEQENPDSSSE 240
QY 241 PVVEDERLHHDTDVTVQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
DB 241 PVVEDERLHHDTDVTVQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
QY 301 FPVEEQOEVPETNRKTDDEPKAKYKKKKPKLLNKFDKTIKAELEDAEKLRRKGKIEEA 360
DB 301 FPVEEQOEVPETNRKTDDEPKAKYKKKKPKLLNKFDKTIKAELEDAEKLRRKGKIEEA 360
QY 361 VNAFKELVRYKYPQSPRARYGKACQEDDLAEKRSNEVLRGAIETQYQEVASLPPVADLLK 420
DB 361 VNAFKELVRYKYPQSPRARYGKACQEDDLAEKRSNEVLRGAIETQYQEVASLPPVADLLK 420
QY 421 LSLKRSDRQQLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDNDNAKKYEEV 480
DB 421 LSLKRSDRQQLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDNDNAKKYEEV 480
QY 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFHLGDAMQVRGN 540
DB 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFHLGDAMQVRGN 540
QY 541 KEAYKMYELGHRGHFASVWQSRSLYVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
DB 541 KEAYKMYELGHRGHFASVWQSRSLYVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLRKGDWSQFTLWQGRRNENACKGAPKCTTLLEKPEFTTGC 660
DB 601 GLAVMDKAKGLFLPEDENLRKGDWSQFTLWQGRRNENACKGAPKCTTLLEKPEFTTGC 660
QY 661 RRGQIKYSIMPGTHVWPHPTGPTNCRMLHGLVLPKEGCKIRCANETRTWEEGKYLIFD 720
DB 661 RRGQIKYSIMPGTHVWPHPTGPTNCRMLHGLVLPKEGCKIRCANETRTWEEGKYLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758
DB 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758

RESULT 3
ADA00639
ID ADA00639 standard; protein; 758 AA.
XX
XX ADA00639;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Human aspartyl (asparaginyl) beta-hydroxylase (HAAH).
DE
XX
XX Tumour growth inhibition; human aspartyl (asparaginyl) beta-hydroxylase;
KW HAAH hydroxylation; NOTCH polypeptide;
KW epidermal growth factor-like repeat; EGF-like repeat; tumour cell;
KW malignant neoplasm; colon cancer; breast cancer; pancreatic cancer;
KW liver cancer; cancer of the bile duct; cancer the central nervous system;
KW CNS; cytostatic; enzyme; human.
XX
XX Homo sapiens.
OS
XX US2003031670-A1.
PN
XX
XX 13-FEB-2003.
PD
XX
XX 08-NOV-1999; 99US-00436184.
PF
XX
XX 08-NOV-1999; 99US-00436184.
PR
XX
XX 08-NOV-1999; 99US-00436184.
XX
XX
XX (WAND/) WANDS J R.
PA (DMON/) DE LA MONTE S M.
PA (INCE/) INCE N.
PA (CARL/) CARLSON R I.
XX
XX
XX Wands JR, De la Monte SM, Ince N, Carlson RI;
PI
XX
XX WPI; 2003-605701/57.
DR
XX
XX N-PSDB; ADA00640.
DR
XX
XX

PT Inhibiting tumor growth or killing tumor cells (e.g. cancer of the colon,
PT breast, pancreatic, liver or the central nervous system), by
PT administering an inhibitor of the human aspartyl (asparaginyl) beta-
PT hydroxylase.

PT Disclosure; Page 2; 30pp; English.

XX
XX
XX The present invention relates to a method for inhibiting tumour growth in
XX a mammal. The method comprises administering to the mammal a compound,
XX which inhibits the expression or enzymatic activity of a human aspartyl
XX (asparaginyl) beta-hydroxylase (HAAH). The compound may inhibit HAAH
XX hydroxylation of a NOTCH polypeptide. In particular, the compound may
XX inhibit hydroxylation of an epidermal growth factor (EGF)-like repeat
XX sequence in a NOTCH polypeptide. The methods are useful for inhibiting
XX tumour growth or killing tumour cells, or for diagnosing or
XX prognosticating a malignant neoplasm. In particular, the tumour or
XX neoplasm is colon cancer, breast cancer, pancreatic cancer, liver cancer,
XX cancer of the bile ducts, or cancer or tumour of the central nervous
XX system (CNS). The present sequence represents HAAH.
XX
SQ Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.1e-311;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGNSSSSGSGSGSSTAGSSSPGARRETKHGKNGRKGLSGTSFFTFMWV 60
DB 1 MAQRKNAKSSGNSSSSGSGSGSSTAGSSSPGARRETKHGKNGRKGLSGTSFFTFMWV 60
QY 61 IALLGWTSAVAVVWFDLVDYEEVGLKIGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
DB 61 IALLGWTSAVAVVWFDLVDYEEVGLKIGIYDADGDGDFVDDAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEPQNIIEDEAKEQIQSLHEWVHAHEVGEDLQOEDGPTGEPQ 180
DB 121 PEEAEPHTEPEEQVPVEAEPQNIIEDEAKEQIQSLHEWVHAHEVGEDLQOEDGPTGEPQ 180

Db 121 PEEAEPHTEPEEQVVEAEFQNIIDEAKEQIQSLIHEMVHAEHVEGEDIQEDGPTGEPO 180
QY 181 QEDDEFIMATDVDRFETLEPEVSHETESHVHEETVSQDCNQMEEEMNSEQENPDSSSE 240
Db 181 QEDDEFIMATDVDRFETLEPEVSHETESHVHEETVSQDCNQMEEEMNSEQENPDSSSE 240
QY 241 PVVEDERLHDDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS 300
Db 241 PVVEDERLHDDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS 300
QY 301 FVVEEQEVPETNRTKTDDEQAKAVKKKKPKLINKEDTIKAEKLDAAEKLKRGKITEEA 360
Db 301 FVVEEQEVPETNRTKTDDEQAKAVKKKKPKLINKEDTIKAEKLDAAEKLKRGKITEEA 360
QY 361 VNAFKEVLVKYPQSPRARVGAQCCEDDLAEKRSNEVLRAIETVQVVASLPDVPADLLK 420
Db 361 VNAFKEVLVKYPQSPRARVGAQCCEDDLAEKRSNEVLRAIETVQVVASLPDVPADLLK 420
QY 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYYLLIGDNDNAKKVYEEV 480
Db 421 LSLKRRSDRQOQFLGHMRGSLTLQRLVQLFPNDTSLKNDLGVGYYLLIGDNDNAKKVYEEV 480
QY 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGREFYHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGREFYHLGDAMQVRGN 540
QY 541 KEAYKMYELGKHKGHFASVWQSLYNVNGLKAQPMWTPKETGYTELVSLENNWKLIRDE 600
Db 541 KEAYKMYELGKHKGHFASVWQSLYNVNGLKAQPMWTPKETGYTELVSLENNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRNENACKGAPKTCULLEKFPETTCG 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRNENACKGAPKTCULLEKFPETTCG 660
QY 661 RRGQIKYSIMHPGTHWPTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
Db 661 RRGQIKYSIMHPGTHWPTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
QY 721 DSFEHEVWQDASSFRLIFIVDVHPELTPQQRSLPAI 758
Db 721 DSFEHEVWQDASSFRLIFIVDVHPELTPQQRSLPAI 758

RESULT 4
ABU92053
ID ABU92053 standard; protein; 265 AA.
XX AC ABU92053;
XX DT 15-JUL-2003 (first entry)
XX DE Human protein modification and maintenance molecule-33 (PMM-33).
XX KW Human; protein modification and maintenance molecule; PMM; cancer;
KW cell proliferation disorder; atherosclerosis; neurological disorder;
KW epilepsy; Huntington's disease; stroke; immune disorder; allergy;
KW inflammatory disorder; AIDS; developmental disorder; hypothyroidism;
KW Cushing's syndrome; gastrointestinal disorder; epithelial disorder;
KW infection; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; anti-allergic; vulnerary;
KW antiinflammatory; thymomimetic.
XX OS Homo sapiens.
XX PN WO2003031939-A2.
XX PD 17-APR-2003.
XX PF 11-OCT-2002; 2002WO-US032850.
XX PR 12-OCT-2001; 2001US-0329689P.
XX PR 25-OCT-2001; 2001US-0335703P.
XX PR 09-NOV-2001; 2001US-0348887P.

PR 28-NOV-2001; 2001US-0334145P.
PR 06-DEC-2001; 2001US-0337451P.
PR 14-DEC-2001; 2001US-0340584P.
XX (INCY-) INCYTE GENOMICS INC.
XX PA Ramkumar J, Gorvad AE, Baughn MR, Emerling BM, Yang J, Lee SY;
PI Tran UK, Becha SD, Duggan BM, Lee EA, Griffin JA, Li JX;
PI Sprague WW, Hafalia AJA, Chawla NK, Lehr-Mason PM, Kable AB, Yue H;
PI Marquis JF, Yao MG, Richardson TW, Tang TY, Jin P, Chien D;
PI Bhatia U, Burrill JD, Lee S, Blake JF, Ho A, Zheng W;
XX WPI; 2003-430274/40.
DR N-PSDB; ACA92448.
XX PT New human protein modification and maintenance molecules (PMM), useful
PT for diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant PMM expression e.g. cancer, atherosclerosis, or
PT infections.
PS Claim 1; Page 273-274; 31pp; English.
XX The present invention relates to the isolation of human protein
CC modification and maintenance molecules (PMM), and the polynucleotide
CC sequences encoding them. A total of 40 PMM polypeptides (designated PMM
CC -1 to PMM-40) are disclosed. The sequences of the invention are useful
CC for diagnosing a condition or disease associated with the expression of
CC PMM in a subject, preparing a polyclonal or monoclonal antibody, and
CC generating an expression profile of a sample containing the
CC polynucleotides. The diseases or conditions associated with decreased
CC expression or overexpression of PMM are cell proliferation disorders
CC (e.g. cancer, atherosclerosis), neurological disorders (e.g. epilepsy,
CC Huntington's disease, stroke), immune/inflammatory disorders, (e.g. AIDS,
CC allergies), developmental disorders (e.g. hypothyroidism, Cushing's
CC syndrome), gastrointestinal or epithelial disorders, and infections. The
CC PMM polypeptides or their fragments are useful in screening compounds
CC for effectiveness as agonists or antagonists of the polypeptides, or in
CC altering the expression of the target polynucleotide and compounds that
CC specifically bind to, or modulate the activity of the polypeptide.
CC ABU92021-ABU92060 represent the human PMM polypeptides of the invention
XX SQ Sequence 265 AA;
Query Match 33.7%; Score 1353.5; DB 6; Length 265;
Best Local Similarity 92.5%; Pred. No. 4.2e-99;
Matches 259; Conservative 2; Mismatches 0; Indels 19; Gaps 1;
QY 34 RETKHGHKNGKRGKLGSTSFPTWEMVIALLGWTSVAVWVFDLVYEEVLGKLGTYDAD 93
Db 5 KETKHGHKNGKRGKLGSTSFPTWEMVIALLGWTSVAVWVFDLVYEEVLGKLGTYDAD 64
QY 94 GDGDFVDVDAKVLILIKERSTSEPAVPPBEAEPHTEPEEQVVEAEFQNIIDEAKEQIQS 153
Db 65 GDGDFVDVDAKVLILIKERSTSEPAVPPBEAEPHTEPEEQVVEAEFQNIIDEAKEQIQS 124
QY 154 LHEMVHAEHVEGEDLQEDGPTGEPQEDDEFIMATDVDRFETLEPEVSHETESHYH 213
Db 125 LHEMVHAEHVEGEDLQEDGPTGEPQEDDEFIMATDVDRFETLEPEVSHETESHYH 184
QY 214 VEETVSQDCNQMEEEMNSEQENPDSSSEPVVEDERLHDDTDVTVQVVEEQAVVEPLENEG 273
Db 185 VEET-----DSSEPVVEDERLHDDTDVTVQVVEEQAVVEPLENEG 225
QY 274 IEITEVTAPEDNPVEDSQVIVEEVSIFVEEQEVPPEP 313
Db 226 IEITEVTAPEDNPVEDSQVIVEEVSIFVEEQEVPPEP 265
RESULT 5
AAY33642
ID AAY33642 standard; protein; 255 AA.
XX AC AAY33642;

XX 06-JAN-2000 (first entry)
DT Human labyrinthin protein.
XX
DE Labyrinthin; human; cancer; marker; antigen; detection; antibody;
KM MCA 44-3A6; diagnostic; vaccine; treatment; adenocarcinoma.
XX
OS Homo sapiens.
XX
PN WO947683-A1.
XX
PD 23-SEP-1999.
XX
PF 11-MAR-1999; 99WO-US005365.
XX
PR 17-MAR-1998; 98US-00040485.
XX
PA (RADO/) RADOSEVICH J A.
XX
PI Radosevich JA;
XX
DR WPI; 1999-580307/49.
DR N-PSDB; AAZ23609.
XX
XX Novel Labyrinthin polynucleotides and polypeptides used as a diagnostic
PT marker for cancer and in anticancer vaccines.
XX
PS Claim 6; Fig 2; 34pp; English.
XX
CC This invention describes a novel polynucleotide and polypeptide sequence
CC for the antigen detected by the antibody MCA 44-3A6. This antigen is
CC designated Labyrinthin (Lab). Antibodies directed to the Labyrinthin
CC (Lab) protein are useful for the detection of Lab. The Lab protein is
CC useful in diagnostic assays for cancer, e.g. to monitor the presence and
CC amount of antibodies (this method is especially useful for cancer cells
CC that have the Lab marker). As the Lab gene is not tissue-specific, it
CC will detect cancer regardless of which organ it occurs in. Peptides
CC derived from Lab are used in the preparation of vaccines to prevent human
CC cancers and/or to treat humans with cancer. Antibody MCA 44-3A6 is able
CC to differentiate antigens associated with adenocarcinomas. However, the
CC sequence of the antigen detected by this antibody has not been elucidated
CC in the prior art. Determination of the polypeptide and polynucleotide
CC sequence of this antigen would enhance its usefulness in cancer
CC diagnosis, treatment and prevention. The present invention discloses the
CC sequences of the antigen recognized by the MCA 44-3A6 antibody. This
CC sequence represents the human lab protein described in the method of the
CC invention
XX
SQ Sequence 255 AA;

Query Match 33.2%; Score 1334; DB 2; Length 255;
Best Local Similarity 99.6%; Pred. No. 1.4e-97;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 59 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db 1 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVEDDAKVLGLKERSTSEPA 60

QY 119 VPPEEAPHTPEBEQVPVEAPQNIIDEAKEQIQSLHEWVHAHVEGEDLQOEDGPTGE 178
Db 61 VPPEEAPHTPEBEQVPVEAPQNIIDEAKEQIQSLHEWVHAHVEGEDLQOEDGPTGE 120

QY 179 PQOEDDEFMATVDVDRFETLEPEVSHETESHYHVEETVSQDCNODMEEMMSEQENPDS 238
Db 121 PQOEDDEFMATVDVDRFETLEPEVSHETESHYHVEETVSQDCNODMEEMMSEQENPDS 180

QY 239 SEPVEDERLHHDTDVTVQVVEEAVYEPLENEGIEITETVTAPEPDNPVEDSQYIVEEV 298
Db 181 SEPVEDERLHHDTDVTVQVVEEAVYEPLENEGIEITETVTAPEPDNPVEDSQYIVEEV 240

QY 299 SIFPVEEQQEVPPET 313
|:|

Db 241 SIFPVEEQQEVPPDT 255

RESULT 6
AAU85544
ID AAU85544 standard; protein; 255 AA.
XX
AC AAU85544;
XX
DT 21-MAY-2002 (first entry)
XX
DE Clone #48005 (L979P) of lung tumour protein.
XX
KM Lung tumour; cancer; T cell; immune response stimulator; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200204514-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001WO-US022058.
XX
PR 11-JUL-2000; 2000US-00614124.
PR 29-AUG-2000; 2000US-00651563.
PR 08-SEP-2000; 2000US-00658824.
PR 26-SEP-2000; 2000US-00671325.
PR 06-OCT-2000; 2000US-00677419.
PR 30-OCT-2000; 2000US-00702705.
PR 13-DEC-2000; 2000US-00736457.
PR 03-MAY-2001; 2001US-00849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
PI Warnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;
PI Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
XX
DR WPI; 2002-164634/21.
DR N-PSDB; ABK39746.
XX
XX Novel polynucleotide encoding a lung tumor polypeptide useful for
PT stimulating and/or expanding T cells specific for a tumor protein.
XX
PS Claim 2; SEQ ID NO 1806; 223pp; English.
XX
CC The invention describes an isolated polynucleotide and polypeptide useful
CC for stimulating and/or expanding T cells specific for a tumour protein
CC for determining the presence of a cancer in a patient. A composition
CC containing the polynucleotide and/or polypeptide is useful for treating a
CC lung cancer in a patient. The polypeptide is useful for removing tumour
CC cells from a biological sample. The polynucleotide is also useful as
CC probe or primer to detect the level of mRNA encoding a tumour protein.
CC This is the amino acid sequence of a lung tumour associated protein,
CC described in the method of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 5; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db 1 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVEDDAKVLGLKERSTSEPA 60

QY 119 VPPEEAPHTPEBEQVPVEAPQNIIDEAKEQIQSLHEWVHAHVEGEDLQOEDGPTGE 178
Db 61 VPPEEAPHTPEBEQVPVEAPQNIIDEAKEQIQSLHEWVHAHVEGEDLQOEDGPTGE 120

QY 179 PQEEDDEFILMATVDVDRFETLEPEVSHSEETSHYVEETVSQDCNQDMSEEMSEQENPDS 238
DB 121 PQEEDDEFILMATVDVDRFETLEPEVSHSEETSHYVEETVSQDCNQDMSEEMSEQENPDS 180
QY 239 SEPVEDERLHNDTDDVTYQVVEEQAVVEPLENEGIEITEVTAPPPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDERLHNDTDDVTYQVVEEQAVVEPLENEGIEITEVTAPPPEDNPVEDSQVIVEEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255

RESULT 7

ABU69516
ID ABU69516 standard; protein; 255 AA.

XX AC ABU69516;

XX DT 05-JUN-2003 (first entry)

XX DE Human lung cancer-associated protein L979P.

XX KW Human; lung cancer; lung tumour; cytostatic; vaccine; T cell expansion;
CD4; CD8.

XX OS Homo sapiens.

XX PN US2002197669-A1.

XX PD 26-DEC-2002.

XX PF 03-MAY-2001; 2001US-00849626.

XX PR 13-DEC-2000; 2000US-00736457.

XX PA (BANG/) BANGUR C S.

PA (FANG/) FANGER G R.

PA (WANG/) WANG A.

PA (WANG/) WANG T.

PA (SWIT/) SWITZER A P.

PA (MCNE/) MCNEILL P D.

PA (CLAP/) CLAPPER J D.

XX PI Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, Mcneill PD;
PI Clapper JD;

XX DR WPI; 2003-352750/33.

XX DR N-PSDB; ACA12072.

XX PS Example 5; Page; 72pp; English.
XX CC The invention relates to a polynucleotide encoding a lung tumour protein,
CC comprising a sequence selected from any of the 14 sequences mentioned in
CC the specification, or a sequence (S2) mentioned in specification,
CC complement of S1, sequences consisting of at least 20 contiguous residues
CC of S1, sequences that hybridise to S1, sequences having 75%, preferably
CC 90%, identity to S1, or degenerate variants of S1. Also included are an
CC isolated polypeptide (comprising a sequence (S3) selected from any one of
CC the 4 amino acid sequences mentioned in the specification, a sequence
CC encoded by the polynucleotide, or sequences having at least 70%,
CC preferably 90%, identity to a sequence encoded by the polynucleotide), an
CC expression vector comprising the polynucleotide operably linked to an
CC the vector, an isolated antibody (or its antigen-binding fragment) that
CC specifically binds to the polypeptide, detecting the presence of a cancer
CC in a patient, a fusion protein comprising the polypeptide, an
CC oligonucleotide that hybridises to S1 under moderately stringent
CC conditions, stimulating and/or expanding T cells specific for a tumour

CC protein (comprising contacting T cells with the polynucleotide, protein
CC or antigen-presenting cells, under conditions and for a time sufficient
CC to permit the stimulation and/or expansion of T cells) and inhibiting the
CC development of a cancer in a patient (by incubating CD4⁺ and/or CD8⁺ T
CC cells isolated from a patient with the polynucleotide, protein or antigen
CC presenting cells that express the polynucleotide, such that T cells
CC proliferate, administering to the patient an effective amount of the
CC proliferated T cells, and thus inhibiting the development of a cancer in
CC the patient. The polynucleotide, protein and cells are useful in a
CC composition for stimulating an immune response in a patient, and for
CC treating a cancer in a patient (particularly lung cancer). The
CC oligonucleotide is useful for determining the presence of a cancer in a
CC patient. The protein and oligonucleotides are useful in pharmaceutical
CC compositions, e.g. vaccines. The polynucleotide is also useful as a probe
CC or primer for nucleic acid hybridisation, and in the design and
CC preparation of ribozyme molecules for inhibiting expression of tumour
CC polypeptides and proteins in tumour cells. An amplified portion of the
CC polynucleotide is useful for isolating a full-length gene from a suitable
CC library. The present sequence is a protein encoded by a cDNA (full
CC length, extended or partial) isolated from a library derived from lung
CC tumour/ cancer cells. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from the USPTO at
CC seqdata.uspto.gov/sequence.html?DocId=20020197669
XX SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 6; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVALLGWTSVAVVWFDLVYEEVLGKLGIDADGGDFVDPAKVLGLKERSTSEPA 118
DB 1 MVALLGWTSVAVVWFDLVYEEVLGKLGIDADGGDFVDPAKVLGLKERSTSEPA 60

QY 119 VPPEEAEPTTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVAHEHVEGEDLQOEDGPTGE 178
DB 61 VPPEEAEPTTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVAHEHVEGEDLQOEDGPTGE 120

QY 179 PQEEDDEFILMATVDVDRFETLEPEVSHSEETSHYVEETVSQDCNQDMSEEMSEQENPDS 238
DB 121 PQEEDDEFILMATVDVDRFETLEPEVSHSEETSHYVEETVSQDCNQDMSEEMSEQENPDS 180

QY 239 SEPVEDERLHNDTDDVTYQVVEEQAVVEPLENEGIEITEVTAPPPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDERLHNDTDDVTYQVVEEQAVVEPLENEGIEITEVTAPPPEDNPVEDSQVIVEEV 240

QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255

RESULT 8
ABU66419
ID ABU66419 standard; protein; 255 AA.

XX AC ABU66419;

XX DT 22-MAY-2003 (first entry)

XX DE Lung cancer therapyand diagnosis associated protein #43.

XX KW Lung cancer; cytostatic; vaccine; gene therapy; cancer.

XX OS Homo sapiens.

XX PN US2002172952-A1.

XX PD 21-NOV-2002.

XX PF 10-JUL-2001; 2001US-00902941.

XX PR 30-JUN-1999; 99US-00346492.

PR	15-OCT-1999;	99US-00419356.
PR	17-DEC-1999;	99US-00466807.
PR	30-DEC-1999;	99US-00476300.
PR	06-MAR-2000;	2000US-00519642.
PR	22-MAR-2000;	2000US-00533077.
PR	10-APR-2000;	2000US-00546259.
PR	27-APR-2000;	2000US-00560406.
PR	05-JUN-2000;	2000US-00589184.
PR	11-JUL-2000;	2000US-00614124.
PR	29-AUG-2000;	2000US-00651563.
PR	08-SEP-2000;	2000US-00658824.
PR	26-SEP-2000;	2000US-00671325.
PR	06-OCT-2000;	2000US-00677419.
PR	30-OCT-2000;	2000US-00702705.
PR	13-DEC-2000;	2000US-00736457.
PR	03-MAY-2001;	2001US-00849626.

PA (CORI-) CORIXA CORP.

PI Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW;
PI Durham M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;
XX
DR WPI; 2003-328427/31.

PT New polynucleotide, useful for preparing a composition for treating or inhibiting development of cancer, e.g. lung cancer.

PS Example 5; SEQID NO 1807; 82pp; English..

CC The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy,
CC for treating or inhibiting development of cancer, e.g. lung cancer. This
CC sequence represents a polypeptide associated with the compositions and
CC methods for the therapy and diagnosis of lung cancer

SQ Sequence 255 AA;

Query Match	32.8%	Score 1320;	DB 6;	Length 255;
Best Local Similarity	98.8%;	Pred. No. 1.9e-96;		
Matches 252; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0;

QY	59	MVIALLGWTSVAVVMFDLVDYBEVLTGLKIGIYDADGDDGDFVDDAKVLLGLKERSTSEPA	118
DB	1	MVIALLGWTSVAVVMFDLVDYBEVLTGLKIGIYDADGDDGDFVDDAKVLLGLKERSTSEPA	60
QY	119	VPPEAEAPHTEBEEQVPVEAEFPONIEDEAKEQIQSLTHEWVAHEVEGEDLQQEDGPTGE	178
DB	61	VPPEAEAPHTEBEEQVPVEAEFPONIEDEAKEQIQSLTHEWVAHEVEGEDLQQEDGPTGE	120
QY	179	PQQEDEFIMATDVDDRFFETLPEVSHHETESHYHVEETVSQDCNQDMBEMMSEQENPDS	238
DB	121	PQQEDEFIMATDVDDRFFETLPEVSHHETESHYHVEETVSQDCNQDMBEMMSEQENPDS	180
QY	239	SEPVEDERLHHDTDVTVQVYEEQAVYERPLENEGIEITEVTAPREDNPVEDSQVIVEEV	298
DB	181	SEPVEDERLHHDTDVTVQVYEEQAVYERPLENEGIEITEVTAPREDNPVEDSQVIVEEV	240
QY	299	SIFPVEEQQEVPPET 313	
DB	241	SIFPVEEQQEVPPDT 255	

RESULT 9
ABB61986
ID ABB61986 standard; protein; 422 AA

AC	ABB61986;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 12750
XX	

XW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS *Drosophila melanogaster*.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EM;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL06089.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
PT interactions.

PS Disclosure; SEQ ID NO 12750; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB12072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 422 AA;
SQ

Query Match	23.2%;	Score 933;	DB 4;	Length 422;
Best Local Similarity	45.0%;	Pred. No. 2.5e-65;		
Matches 184;	Conservative 75;	Mismatches 144;	Indels 6;	Gaps 5;

QY	355	GKIEEAVNAFKELVRKYPQSPRÄRYGKAÖCEDDLAEKRSNEVLRGAIETTYÖEAVSLPDV	414
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	15	GNYAÖALRSFNTLTTNFAPHERSAHLGRARLLELLAKKERSNQRLWEAIDAYKRYLAÆGEL	74
QY	415	PADLTKLSLKRRS--DRÖQFLGMRGSLTLÖRLVÖLPFNDTSLKNDLGVGYLLIGDNDN	472
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	75	VÄSNÖEFQTAGESCIENTLRFLGHRÖATTIHELLINRLPEDRLRNÖLSTLYLMVNNLÖQ	134
QY	473	AKKVYEEVLSVTPNDFAKVHYFLKA-ÖNKIAESIPLYKEGIESGDPGTDDGREYFHL	531
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	135	VEKVAVELTKLMPNNAVÄÖLHYGLALRÖFHADYAKALPYLYKVAVESGEGTÖEAFYLSL	194
QY	532	GDAMÖRVGNK-EÄYKMYELGHRGHFASVWÖRSLYNNVNLKÄÖPMWTPKETGYTELVSXL	590
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	195	GETMÖRLMSKSEÄLEVYKGVAKGFÄSLYÖRSLYNEPRLRÄÖPFWÖPKETGYERÖEKL	254
QY	591	ERNWKLTRDEGLAVMDKAKGLFLPEDENLREKGDWSÖFTLWÖÖGRRENENACKGAPKTCTL	650
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	255	TLNWRÄIRDEGLALLGRS-GFFEDEÄELLRDKGWVÖQYELYÄÖGRRVKDNCRÄPRTCSL	313
QY	651	LEKFPETTGCRRGÖIKYSIMHPGITHVPHTGPTNCRRLRMHLGLVIPK-EGCKIRCANETR	709
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	314	LEEPFESAGCRRGÖVKFSVMÖAKITHVPHCGPTNCRRLRAHLTLAÄPEPEKASLRVÄÖGR	373
QY	710	TWEEGKVLTFDDSFHEHEVWÖDASSFRLLFIVDVMHPELTPÖÖRSLPAI	758
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	374	TWREGELTFDDSFHEHEVWANGÖSRVLVLLDMWHPRÖSÄÄÖRSLSPI	422

RESULT 10	
AAB73682	
ID	AAB73682 standard; protein, 369 AA.
XX	
AC	AAB73682;
XX	
DT	11-SEP-2001 (first entry)
XX	
DE	Human oxidoreductase protein ORP-15.
XX	
KM	Human oxidoreductase protein; ORP; cell proliferative disorder;
KM	arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;
KM	diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhoea;
KM	osteoporosis; metabolic disorder; obesity; phenylketonuria;
KM	hypercholesterolaemia; reproductive disorder; infertility;
KM	ovulatory defect; menstrual cycle defect; endometriosis; chromosome 22;
KM	polycystic ovary disease; spermatogenesis disruption; impotence;
KM	neurological disorder; epilepsy; stroke; Alzheimer's disease;
KM	Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;
KM	meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;
KM	schizophrenic disorder; infection; autoimmune disorder;
KM	inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;
KM	allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;
KM	rheumatoid arthritis; ulcerative colitis; drug screening;
KM	toxicity screening; transgenic animal; SNP detection; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200144448-A2.
XX	
XX	21-JUN-2001.
XX	
PF	07-DEC-2000; 2000WO-US033158.
XX	
XX	16-DEC-1999; 99US-0172367P.
PR	
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DAM;
XX	
DR	WPI; 2001-390245/41.
XX	
DR	N-PSDB; AAH24237.
XX	
PT	Novel human oxidoreductase protein (ORP) useful for diagnosing, treating
PT	and preventing cell proliferative, neurological, viral, reproductive and
PT	autoimmune/inflammatory disorders associated with abnormal expression of
PT	ORP.
XX	
PS	Claim 1; Page 110-111; 136pp; English.
XX	
CC	Sequences AAB73688-AAB73694 represent 27 novel human oxidoreductase
CC	proteins, designated ORP-1 to ORP-27 respectively, and sequences AAH24223
CC	-AAH24249 represent cDNAs encoding ORP-1 to ORP-27. Human ORP proteins
CC	and nucleic acids are useful for diagnosing, treating or preventing cell
CC	proliferative disorders (e.g. arteriosclerosis, cirrhosis, psoriasis,
CC	cancers); endocrine disorders (e.g., type I or II diabetes mellitus,
CC	diabetes insipidus, dwarfism, hirsutism, amenorrhoea, osteoporosis);
CC	metabolic disorders (e.g., obesity, phenylketonuria,
CC	hypercholesterolaemia); reproductive disorders (e.g., infertility,
CC	ovulatory and menstrual cycle defects, endometriosis, polycystic ovary
CC	disease, disruption of spermatogenesis, impotence); neurological
CC	disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's
CC	disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease,
CC	cerebral palsy, muscular dystrophy, mood, anxiety and schizophrenic
CC	disorders); viral, bacterial, fungal and parasitic infections; and
CC	autoimmune/inflammatory disorders such as acquired immunodeficiency
CC	syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis,
CC	gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis.
CC	Human ORP proteins and nucleotides can be used to identify compounds
CC	which modulate their activity or expression. ORP nucleic acid sequences
CC	may also be used for assessing the toxicity of a test compound, to detect
CC	upstream sequences such as promoters and regulatory elements, and to
CC	create knock out or knock in animals or transgenic animals to model human

[illegible]

XX Shimkets RA, Leach M;
PI
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC77536.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 11, Page 5365; 5507bp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 236 AA;
Query Match 6.8%; Score 271.5; DB 3; Length 236;
Best Local Similarity 31.2%; Pred. No. 4.7e-13;
Matches 72; Conservative 40; Mismatches 100; Indels 19; Gaps 8;
QY 544 YKMYELG-HK---RGHF---ASVQSRSLYVNGLKAQPMWTPKETGYTELKSLERN 593
Db 4 YSMGSMGRHKIGREQGRYLNSRPSIQKPEVFLPDLPTPYFSRDAQKHD--VEVLERN 61
QY 594 WKLI RDEGLAVMDKAKGLFLPE--DENLREKGDWSQFTLWQGRNENACKGAPKTCITLL 651
Db 62 FQTLICEFETLVKAFNSNCSLPQGWKXNSTRSGEWFTEFLVNGVCVPRNCRKCPRTYLL 121
QY 652 EKFPETTGCR-RGQIKYSIMHPTGTHVPHPTNCRRLRMHLGLVIPKEGCKIRCANETRT 710
Db 122 GSLRTICGNNVFENACISVLSPTGVTITEHYGPTNIRIRCHLGLKTP-NGCELVVGGEPOC 180
QY 711 WEEGKVLIFDDSFHEHVMQDASSF--RLIFIVDVWHPPELTPOQRSLPAI 758
Db 181 WAEGRCLLFDDSFHLHAFHEGSAEDGPRVVFVVDLMHPNVAAAEERQALDFI 231
RESULT 12
ABM68322
ID ABM68322 standard; protein; 264 AA.
XX
AC ABM68322;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #1419.
XX
KM Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KM detection; food; gene expression; plant; animal; microorganism; toxin;
KM antibiotic; biopesticide; virulence factor; disease model; plague;
KM whooping cough.
XX
OS Photorhabdus luminescens.
XX

PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
DR
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 1419; 1205bp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 264 AA;
Query Match 6.5%; Score 262; DB 6; Length 264;
Best Local Similarity 29.1%; Pred. No. 3.2e-12;
Matches 66; Conservative 39; Mismatches 114; Indels 8; Gaps 3;
QY 533 DAMQVRGNKEAYKMYELGHKRGHPSVQSRSLYVNGLKAQPMW-TPKETGYTELKSL 591
Db 35 DCLKRVED---MFPPLQQR--APLQADAKYIMPGISTPWLDTNSFPQLOPLVTSIM 87
QY 592 RNWKLIRDEGLAVMDKAKGLFLPEENLREKGDWSQFTLWQGRNENACKGAPKTCITLL 651
Db 88 NNADKIKQETNAVVISGESQYITDYEHYLGTDKDWKALYLFKNGQPNNAVANILPATWITF 147
QY 652 EKFPETTGCRGQIKYSIMHPTGTHVPHPTNCRRLRMHLGLVIPKEGCKIRCANETRTW 711
Db 148 NNELRDWHCPLELVHFSVLQPGTVIKPHCDLWNFTLNLHFAVDIPASHCEIIVANEARCW 207
QY 712 EEGKVLIFDDSFHEHVMQDASSFRLIFVDVWHPPELTPOQRSLPAI 758
Db 208 KEGECLEFDYSQHEAYNRSKRIICLLMDIWHPNLSFAERBALVLI 254
RESULT 13
AAU29679
ID AAU29679 standard; protein; 104 AA.
XX
AC AAU29679;
XX
DT 18-DEC-2001 (first entry)
XX

Db 74 IRDEAFALRDASHIRATAYNDIGFNSFFRNGMRFFLYLKWYGRPHPSAVALCPTVELLG 133

QY 653 KEPTTGCRRGOIKYSIMHPGTHWPHPTGPTNCRMLGLVIRK-EGCKIRCANETRTW 711

Db 134 RIPSVKAAAMFAQLP-----PGRLGLHRDPYAGALRYHLGLATPNHDGCAIVDGETYSW 188

QY 712 EBGKYLIFDPSFEHEVWQDASSFRLIFIVDWHP 745

Db 189 RDGEDIVFDETYLHYAFNDTQEDRIILFCDIERP 222

RESULT 15

AAU28081

ID AAU28081 standard; protein; 324 AA.

XX

AC AAU28081;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secretory protein, Seq ID No 250.

XX

KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock; ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;

KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn; amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;

KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease; gut protection; lung; liver fibrosis; immune deficiency; infection;

KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy; multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma; fertility; analgesic; pain; antigen.

XX

OS Homo sapiens.

XX

PN WO200166689-A2.

XX

PD 13-SEP-2001.

XX

PF 05-MAR-2001; 2001WO-US004942.

XX

PR 07-MAR-2000; 2000US-00519705.

PR 19-MAY-2000; 2000US-00574454.

PR 17-JUN-2000; 2000US-00596193.

PR 14-JUL-2000; 2000US-00616847.

PR 19-SEP-2000; 2000US-00665363.

PR 20-OCT-2000; 2000US-00693267.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou F;

PI Zhao QA, Yang Y, Dmanac RT, Zhang J, Chen R, Chen R, Xue AJ, Wang J;

XX

DR WPI; 2001-589934/66.

DR N-PSDB; AAS44981.

XX

PT Novel polypeptides and nucleic acids obtained from cDNA libraries

PT prepared from various human tissues, for diagnosis and treatment of

PT cancer, neurological, inflammatory, and autoimmune disorders.

XX

PS Example 4; SEQ ID NO 250; 107bp; English.

XX

CC The invention relates to novel isolated human secreted polypeptides (I)

CC and polynucleotides (II). (I) and (II) are useful for treating

CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,

CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is

CC involved in increasing haematopoiesis, stem cell survival, bone growth

CC and remodeling. (I), (II) and modulators of (II) are useful for

CC prophylaxis or treatment of one or more cancers. (II) is also useful for

CC creating transgenic animals useful for studying the in vivo activities of

CC the polypeptide as well as for studying modulators of the polypeptides.

CC (I) induces the proliferation of neural cells and regeneration of nerve

CC and brain tissue and is useful for the treatment of central and

CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,

CC Parkinson's disease, Huntington's disease, and amyotrophic lateral

CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic

CC activity, regulation of haematopoiesis and is useful for treating myeloid

CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia

CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve

CC tissue growth, and in tissue repair, healing of burns, incisions, ulcers,

CC for treating osteoporosis, osteoarthritis, bone degenerative disorders,

CC or periodontal disease. Furthermore, (I) is also useful for gut

CC protection or regeneration and treatment of lung or liver fibrosis, and

CC reperfusion injury in various tissues, various immune deficiencies and

CC disorders including severe combined immunodeficiency (SCID), bacterial or

CC fungal infections, autoimmune disorders e.g. multiple sclerosis,

CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic

CC reactions and conditions, such as asthma or other respiratory problems.

CC In addition, (I) affects biorhythms or circadian cycles of rhythms,

CC fertility, metabolism, catabolism, anabolism, storage or elimination of

CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides

CC analgesic effects or other pain reducing effects, immunoglobulin like

CC activity and can act as an antigen in a vaccine composition to raise an

CC immune response. AAU28020-AAU28395 represent novel human secreted protein

CC amino acid sequences of the invention

XX

SQ Sequence 324 AA;

Query Match 5.4%; Score 217.5; DB 4; Length 324;

Best Local Similarity 36.6%; Pred. No. 1.5e-08;

Matches 53; Conservative 19; Mismatches 56; Indels 17; Gaps 5;

QY 625 WS-----QFTWQQR--RNENACKGAPKCTTLLEKFPETGCRGQIKYSIMHPG 673

Db 175 WSPPLAPGCYQLLYQAGRCQPSNCRCPGAYRALRGLRSFMSAN--TFGNAGFSVLLPG 232

QY 674 THWPHGTPTNCRMLGLVIRK-EGCKIRCANETRTWEKGKYLIFDPSFEHEVWQDAS- 732

Db 233 ARLEGRCGPTNARVRCHLGLKIP-PCGCELVVGEPQCWABEGHCLLVDSFLHTVAHNGSP 291

QY 733 --SFRLIFIVDWHPELTPQQRSL 755

Db 292 EDGPRVVFIVDLMHPNVAGAERQAL 316

Search completed: May 19, 2004, 15:57:47

Job time : 99.9924 secs


```
Db      61  VPPPEAAPHTEPEEQVPEAEFQNIIDEAKQIQSLHMHVHAHEHVEGEDLQQEDGPTGE 120
QY      179  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 238
Db      121  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 180
QY      239  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db      181  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 240
QY      299  SIFPVEEQEQEVPPEP 313
Db      241  SIFPVEEQEQEVPPEP 255
```

```
RESULT 2
US-09-702-705-1806
; Sequence 1806, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-702-705-1806
```

```
Query Match      32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59  MVTALLGWTSVAVVWFMDLVYEEVLGKGIYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db      1  MVTALLGWTSVAVVWFMDLVYEEVLGKGIYDADGDFDVEDDAKVLGLKERSTSEPA 60
QY      119  VPPPEAAPHTEPEEQVPEAEFQNIIDEAKQIQSLHMHVHAHEHVEGEDLQQEDGPTGE 178
Db      61  VPPPEAAPHTEPEEQVPEAEFQNIIDEAKQIQSLHMHVHAHEHVEGEDLQQEDGPTGE 120
QY      179  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 238
Db      121  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 180
QY      239  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db      181  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 240
QY      299  SIFPVEEQEQEVPPEP 313
Db      241  SIFPVEEQEQEVPPEP 255
```

```
RESULT 3
US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1806
```

```
Query Match      32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59  MVTALLGWTSVAVVWFMDLVYEEVLGKGIYDADGDFDVEDDAKVLGLKERSTSEPA 118
Db      1  MVTALLGWTSVAVVWFMDLVYEEVLGKGIYDADGDFDVEDDAKVLGLKERSTSEPA 60
QY      119  VPPPEAAPHTEPEEQVPEAEFQNIIDEAKQIQSLHMHVHAHEHVEGEDLQQEDGPTGE 178
Db      61  VPPPEAAPHTEPEEQVPEAEFQNIIDEAKQIQSLHMHVHAHEHVEGEDLQQEDGPTGE 120
QY      179  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 238
Db      121  PQQEDDEFLMATDVDDRFETLEPEVSHEETESHVVEETVSQDCNQDMEMMSEQENPDS 180
QY      239  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 298
Db      181  SEPVEDERLHMDTDVTVQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEV 240
QY      299  SIFPVEEQEQEVPPEP 313
Db      241  SIFPVEEQEQEVPPEP 255
```

```
RESULT 4
US-09-671-325-1806
; Sequence 1806, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-671-325-1806

Query Match	32.8%	Score 1320;	DB 4;	Length 255;
Best Local Similarity	98.8%	Pred. No. 2,1e-99;		
Matches 252; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0;

QY	59	MVIALLGWMTSVAVVWPFDLVDYEEVLGKLGITYDADGGDFVDVDAKVLGLKERSTSEPA	118
Db	1	MVIALLGWMTSVAVVWPFDLVDYEEVLGKLGITYDADGGDFVDVDAKVLGLKERSTSEPA	60
QY	119	VPPEEAPHTPEEEQVPVEAEFPONIDEAKEQIQSLHHEWYAHAEHVEGEDLQOEDGPTGE	178
Db	61	VPPEEAPHTPEEEQVPVEAEFPONIDEAKEQIQSLHHEWYAHAEHVEGEDLQOEDGPTGE	120
QY	179	PQOEDEFLMATDVDDRFETLLEPEVSHETESHSYHVEETVSCDNQDMEEHMSQENPDS	238
Db	121	PQOEDEFLMATDVDDRFETLLEPEVSHETESHSYHVEETVSCDNQDMEEHMSQENPDS	180
QY	239	SEPVVEDERLHHDTDVTTYQVYEEQAVYERPLENEGIEITETVTAAPEDNPVEDSQVYVEEV	298
Db	181	SEPVVEDERLHHDTDVTTYQVYEEQAVYERPLENEGIEITETVTAAPEDNPVEDSQVYVEEV	240
QY	299	SIFPVEEQQSEVPPEET 313	
Db	241	SIFPVEEQQSEVPPEET 255	

RESULT 5
5231168-2
; Patent No. 5231168
; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN,
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 2:
; LENGTH: 783
5231168-2

Query Match	4.9%;	Score 199;	DB 6;	Length 783;
Best Local Similarity	20.6%;	Pred. No. 1.6e-07;		
Matches 143;	Conservative 134;	Mismatches 237;	Indels 180;	Gaps 36;

[illegible]

```

QY      423 LKRSDRQÓFLGHMRGSLTLQRLVÓLPNDTSLKNDLG-----VGYLLIGDNDNAKV 476
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      504 EILPEDKNEKVQH-----EIVEVEIILPED--KNEKGQHEIVEVEEILPEDKNEKGQ 554
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      477 YE--EVLSTPNDGFAKVHYGFI-----LKAQNKIAESIPLYKEGIESGDPGTD 523
      : | : | : | : | : | : : : : : : : : : : : : : : : : :
Db      555 HEIVEVEEILPEDKNEKVQHEIVEVEEILPEDKNEKVQHEIVEVEEILPEIYE----- 607
      : | : | : | : | : | : : : : : : : : : : : : : : : : :
QY      524 DGRFYFHLGDAMQVRVNKNEAYKMYELGKRGHFASVMQRLYN--VNGLKAQPMWTPK- 579
      : : : : : | : | : : : : : : : : : : : : : : : : :
Db      608 -----IEBVPSTNNNNENIETIKPEEKNEF-SVEEKAIQPEPVVPTLNENENVTPKP 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      580 ---ETGYTELKSLERANKLIRD---EGLAVMDKAKGL--FLPEDENLREKGDMSQFT 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      660 SEGESTIKPDIVQI-----KIVQENKPNKKEIPLVVDGEPKHVEQNIQEDDNDDEDDIDIFE 714
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      630 LW-----QGRRNENACKGAPKCTLLEKPEPTT 658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      715 GLSRDDEKDSNNKNNKKSSPIYISTKKFKKVS 748

```

```

RESULT 6
US-09-252-991A-29355
; Sequence 29355, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29355
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29355

```

Query Match	4.9%;	Score 197.5;	DB 4;	Length 320;
Best Local Similarity	25.9%;	Pred. No. 5.7e-08;		
Matches	58;	Conservative	34;	Mismatches 85;
			Indels	47;
			Gaps	9;

```

QY      551 HKRGHTASVWQSRSLYNNGLKAQPMWTPKETGYTEL-----VKSL 590
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      31 HFRGARLPFLRQLVNHSA-----WFAF---YNSLMYLFSSVPSKPYLDRSRFEDEL 81
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      591 ERNWKLI RDEGLAVMDKAKGLFLPEDENLREKG-----DWSQFTL-WQOGRNENACK 642
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      82 KNNWQTI REEALNLFDEG--YIRALNNNEAGFGSFFKKGWKRFYLLTWYDGPL-PSAQO 137
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      643 GAPKCTLLEKFPETTGCRGQIKYSIMHPGTHVPEHTPTNCRLRMHLGLVIP-KEGCK 701
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      138 LCPKTELVSRIPNVKA-----MFTLLPGSHLNPHRDPFGSLRYHLGLSTPNSDNCR 192
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      702 IRCANETRTWEEGKVLIFDSSFHEHVVQDASSFRLLFIVDVWHP 745
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      193 IYVDGQPYAWRDGEDVDMEDETFVHWKNETEQTIVLLFCIDIERP 236
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2

```


[illegible]

```

RESULT 9
; US-09-092-458-2
; Sequence 2, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES

```

; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium vivax
 ; IMMEDIATE SOURCE:
 ; CLONE: PVM3.3.1
 ;
 US-09-092-458-2

Query Match	4.98;	Score 197.5;	DB 3;	Length 1018;
Best Local Similarity	23.2%;	Pred. No. 3.1e-07;		
Matches 128;	Conservative 86;	Mismatches 205;	Indels 133;	Gaps 27;

```

QY 109 LKERSTSEPAVPPEEABP-----HTEPEEQVPVEAEQONIED-----BAKEQIQSL 155
Db 533 LKPDAGEAVTVPSKEAPVQVPVAVGPAQOEVPTEELMQOEHDDELEGTAEAPPEGELVL 592
QY 156 HEMVHAHVEGEDLQOEEDGP-TGEPOQED--DEFIMATDVDDRFEETLEPEVSHEETHSY 212
Db 593 -----EGEGEPTEEPREGEPTEGEVPEEELATPEDD-FELEEP--TGEVEEETV 640
QY 213 HVEETVSQDCNQDMEMMSQENPDSSSEPVVEDERLHHTDDVTYQVVEBQAVYEPLENE 272
Db 641 EGEETAEE---GEEVEEVPAAVE-----EVEEVPAAVEEVEEVEEVEEVEV 681
QY 273 GIEITVTAPEPDNPVEDSQVIVEVSIFFVEEQOEVPBETNRKTTDDPEQAKAVKKKKPK 332
Db 682 PAEEVEEVEEVPPEE-----VEEVP---EEVEEVPPEEVEEVPPEEVEEVEEVEVP 728
QY 333 LNKFDKTIKAFELDAEKLKRGKIEEAVNAFKEIVKYPQSPRARYGKAQCEDDLAEKR 392
Db 729 AVVEVEVPAAVEEVPPEEVEEVEEVEEVEEVEEDVLQVLPSEEDIQDKPK-KDELGSGI 787
QY 393 RS-----NEVLGAIETYQEVASLPDVPADLLKLSLKRSDROQFLGMRGSLTLQRL 446
Db 788 LSIIDMHYQVKEKEMEHEEBETAVPPLKPEDF---AKEDSQSTEMLTFIQGLEGDWERL 843
QY 447 VOLFPNDTSL-----KNDLGVGYLLIGDNDNAKKVYEBVLSVTPND--GFAKVHY 494
Db 844 -----EVSINKARBRWMEQRNKEMAGWLRLIENKWE--YSQI-STKGKDPAGLRKREW 894
QY 495 G-----FILKAQNKIAIESIPYLKEGIESGDPGTDDGRFFYFHLGDAMORVGNKAEAYKY 547
Db 895 SDEKWKWKWFAEVKSQIDS--HLKKWMND---THSNLFKILVKD-MSQFENKKTKEWL 946
QY 548 -----ELGHRKGF-----ASVQORSLYNVNGLKAQ--PWWTPKETGY-- 583
Db 947 MNHWKKNERGYGSESFEVMTTSKLLNVAKSGREWRANPNINRERRELMKWFLKENEYLG 1006
QY 584 --TELVKSLERN 593
Db 1007 QRMKMDSLEKS 1018

```

RESULT 10
 US-08-728-323A-2
 ; Sequence 2, Application US/08728323A
 ; Patent No. 5948676
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
 ; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
 ; TITLE OF INVENTION: Encoding Same And Uses Thereof
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.

QY 356 KIEAVNAFKELVRKYPQ----SPRAYGKAQCCEDDLAEKRNRNEVLRAIETTYQEVASL 411
Db 913 ELEVEEQEQGVQEQEQETVEEPIIHGSSS-EDEM-----EVDYPVSTHQAASS 964
QY 412 PDVPAD 417
Db 965 P--PGD 968

RESULT 13

US-09-328-352-8015
; Sequence 8015, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8015
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8015

Query Match 4.7%; Score 189.5; DB 4; Length 312;
Best Local Similarity 27.8%; Pred. No. 2.5e-07;
Matches 52; Conservative 34; Mismatches 88; Indels 13; Gaps 4;

QY 564 LYNVNGIKAQFWMTPKETGYTELIVKSLERNMKLIRDEGLAVMD---KAKGFLPEDENL 619
Db 53 MYIFSKVNPQPY---IDTQHFQDLKVLIDENWEMIRDEAKALYDQGIIKASSTYNDLGFNS 109
QY 620 REKGDMSQFTLWQGRNRNENACGAPKCTTLLEKFEETGCRGQIKYSIMHPTGHWPH 679
Db 110 FFKTGMKRFYLYKWESSHPSAELCPKTTALKTLPYTKAA-----MFTELAPDSRLVRH 164
QY 680 TGPTRCRLRMHLGLVTPKEG-CKIRCANETRTWEEGKVLIFDDSFHEHVMQDASSFRLIF 738
Db 165 RDPYAGSLRHLGLITPNDRCRFDVDGERYSWRDQGVFDETYIHYAENKTDQNRITF 224
QY 739 TVDVWHP 745
Db 225 FADVERP 231

RESULT 14

US-08-961-083-218
; Sequence 218, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-218

Query Match 4.4%; Score 177.5; DB 3; Length 565;
Best Local Similarity 22.5%; Pred. No. 5.5e-06;
Matches 114; Conservative 71; Mismatches 199; Indels 123; Gaps 22;

QY 111 ERSTSEPAVPPEE-ABPHTPEEQVPVEAPQNIIDEAKQIQSLHMYHAHVEGED 168
Db 62 KEDSAEPA-PVEEVGGEVSKPEKVAVKESQSDKPAESKVEQAGEPV---APRED 116
QY 169 LQGEDGPTGEPQDEDEFIMATDVDDRFETLEPEVSHETESHYHVEETVSQDCNQ-DME 227
Db 117 EKAPVEPEKQPEAPPEE--KAVEETPKQESTPDTKAETVEP--KEETVNOISIQPKVE 172
QY 228 EMNSEQENPDSSPEVVEDE---RLHHTDVTYQVYEEQAVVEPLENEGELITEVTAP 283
Db 173 TPAVEKQTEPTTEEPKVEQAGEPVAPREDEQAPTAPEPEKQPEVPEEKAV--EETPKP 229
QY 284 ED-----NPVEDSQV---IVEVSIPEVE-----BQQ 307
Db 230 EDKIKIGITKEPVDKSELNNQIDKASSVSPDYSTASYNALGPVLETAKGVYASEPVKOP 289
QY 308 EVPEETNR-----KTD-----DPEQAKVKKKKPKLTKKFDKTIKAELEDAEK 350
Db 290 EVNSETNKLKTAIDALNVDKTELNNITADAKTKVKEHYSRDSWQNLQTEVTKAEKVAANT 349
QY 351 LRKRGKIEAVNAF-----KELVRKYPQSPRAYGKAQ 383
Db 350 DAKQSEVNEAVEKLTATIEKVELSEKPIILTLSTDKKILEREAVAKYTL---NQNKTK 406
QY 384 CEDDLAEKRNRNEVLRAIETQEVASLPVPADLLKLSLKRSDRQQLGMRGSLTL 443
Db 407 IKSITAEKKGEVINTVLTDDKVT-ETISAAFKNLEYYKEYTLSTMTYDRNGEET 465
QY 444 QRL---VOLFPNDTSLKN---DLGVGYLLIGDNDNAKKVVEVLSTPNDGFAKVHYG 495
Db 466 ETLLENQNIQLDKKVELKNIKRTDL-IKY-----ENGETNESLITTIIPDD---KSNYY 515
QY 496 FILKAQNKIAE--SIPYLKEGIESGDP 520
Db 516 LKITSNNQKTTLLAVAKNIETTVNGTP 542

RESULT 15

US-09-536-784-218
; Sequence 218, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-09-536-784-218

```

```

Query Match      4.4%; Score 177.5; DB 4; Length 565;
Best Local Similarity 22.5%; Pred. No. 5.5e-06;
Matches 114; Conservative 71; Mismatches 199; Indels 123; Gaps 22;

QY 111 ERSTSEPAVPPEE--AEPHTEPEEQVPVPAEPQNIIEDEAKEQIOSLHEMVHAHEVGED 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 KEDSAEPA-PVEEVGGEVESKPEEKVAVKESQPSDKPAEESKVEQAGEPV---APRED 116

QY 169 LQEDGPTGEPQDEDEFLMATDVDDRFETLEPEVSHETESHYVEETVSQDCNQ-DME 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 EKAPVEPEKQPEAPEEE--KAVEETPKQESTPDTKAETVER--KETVNQSIQPKVE 172

QY 228 EMSEQENPDSSSEPVVEDE---RLHHTDDVTYQVYEEQAVYEPLENEGIEITEVTAP 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 TPAVEKQTEPTEPEKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEEEKAV---EETPKP 229

QY 284 ED-----NPVEDSQV---IVEEVSIFPVE-----EQQ 307
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 EDKIKGIGTKPEPVDSKSELNNQIDKASSVPTDYSTASYNALGPVLETAKGVYASEPVKQ 289

QY 308 EVPPETNR-----KTD-----DPEQKAKVKKKKPKLNKFKDTIKAEIDAAEK 350
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 EVNSETNKLKTAIDALNDVKTETLNNTIADAKTKVKEHYSDRSQNLQTEVTKAKEKVAANT 349

QY 351 LKRGKIEEAVNAF-----KEIVRYKYPQSPRARVYGAQ 383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 DAKQSEVNEAVEKLTATIEKVELSEKPLTLTSTDKILERAVALYTLTLE---NQNKT 406

QY 384 CEDDLAEKRRSNEVLRGALETYQEVASLPDVPADLLKLSLKRSDRQQLGHMRGSLTL 443
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 IKSITAEIKGEEVINTVLTDDKVTT-ETISAPKMLEYKEYTTLSTMTIYDRNGEET 465

QY 444 QRL---VOLFPNDTSLKN---DLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKYHYG 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 ETLNQNQLDLKKVELKNIKRTDL-IKY-----ENGKETNESLITTPDD---KSNYY 515

QY 496 FILKAQNKIAE--SIPYLKEGIESGDP 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 LKITSNNQKTTLLAVKNIEFTVNGTP 542

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:59:55 ; Search time 74.9354 Seconds

(without alignments)
2814.722 Million cell updates/sec

Title: US-09-903-199-2

Perfect score: 4022

Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVHPELTPOQRSLPAI 758

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 segs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4022	100.0	758	9	US-09-903-248-2	Sequence 2, Appli
2	4022	100.0	758	9	US-09-859-604-2	Sequence 2, Appli
3	4022	100.0	758	9	US-09-903-063-2	Sequence 2, Appli
4	4022	100.0	758	9	US-09-903-216-2	Sequence 2, Appli
5	4022	100.0	758	9	US-09-903-199-2	Sequence 2, Appli
6	4022	100.0	758	9	US-09-903-023-2	Sequence 2, Appli
7	4022	100.0	758	10	US-09-436-184-2	Sequence 2, Appli
8	1320	32.8	255	9	US-09-736-457-1806	Sequence 1806, Ap
9	1320	32.8	255	9	US-09-902-941-1806	Sequence 1806, Ap
10	1320	32.8	255	9	US-09-849-626-1806	Sequence 1806, Ap
11	1320	32.8	255	12	US-10-283-017-1806	Sequence 1806, Ap
12	1320	32.8	255	14	US-10-017-754-1806	Sequence 1806, Ap
13	1320	32.8	255	14	US-10-113-872-1806	Sequence 1806, Ap
14	275	6.8	250	14	US-10-156-761-8153	Sequence 8153, Ap
15	274.5	6.8	369	14	US-10-168-274-15	Sequence 15, Appli

16	218.5	5.4	299	12	US-10-282-122A-49023	Sequence 49023, A
17	217.5	5.4	324	12	US-10-221-278-250	Sequence 250, App
18	217.5	5.4	324	15	US-10-291-172-250	Sequence 250, App
19	199	4.9	109	14	US-10-168-274-5	Sequence 5, Appli
20	196.5	4.9	355	12	US-10-282-122A-51120	Sequence 51120, A
21	190.5	4.7	1162	12	US-09-894-273-2	Sequence 2, Appli
22	190.5	4.7	1162	14	US-10-294-804-2	Sequence 2, Appli
23	189.5	4.7	304	12	US-10-282-122A-45188	Sequence 45188, A
24	189.5	4.7	1616	10	US-09-820-843A-16	Sequence 16, Appli
25	189.5	4.7	1616	12	US-10-282-122A-63593	Sequence 63593, A
26	186	4.6	50	16	US-10-443-622-48	Sequence 48, Appli
27	182	4.5	350	13	US-10-101-487-58	Sequence 58, Appli
28	177.5	4.4	565	9	US-09-765-272-218	Sequence 218, App
29	177	4.4	240	13	US-10-101-487-75	Sequence 75, Appli
30	177	4.4	845	12	US-10-205-331-61	Sequence 61, Appli
31	175	4.4	1000	14	US-10-128-714-3305	Sequence 3305, Ap
32	174	4.3	566	12	US-10-424-599-172054	Sequence 172054, A
33	172.5	4.3	1302	12	US-10-282-122A-46112	Sequence 46112, A
34	170	4.2	299	12	US-10-282-122A-50678	Sequence 50678, A
35	168.5	4.2	200	13	US-10-101-487-53	Sequence 53, Appli
36	168.5	4.2	300	12	US-10-282-122A-60256	Sequence 60256, A
37	168	4.2	302	12	US-10-282-122A-75844	Sequence 75844, A
38	167.5	4.2	1876	12	US-10-282-122A-73861	Sequence 73861, A
39	167	4.2	187	13	US-10-101-487-50	Sequence 50, Appli
40	167	4.2	302	9	US-09-815-242-13815	Sequence 13815, A
41	165.5	4.1	1095	14	US-10-128-714-8305	Sequence 8305, Ap
42	165	4.1	518	12	US-10-425-114-71472	Sequence 71472, A
43	164	4.1	1957	15	US-10-369-493-2070	Sequence 2070, Ap
44	163.5	4.1	1297	12	US-10-282-122A-71692	Sequence 71692, A
45	163.5	4.1	1501	9	US-09-924-154-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1

US-09-903-248-2

Sequence 2, Application US/09903248

Patent No. US20020102263A1

GENERAL INFORMATION:

APPLICANT: Wanda, Jack R.

APPLICANT: de la Monte, Suzanne M.

APPLICANT: Ince, Nedim

APPLICANT: Carlson, Rolfe I.

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS

FILE REFERENCE: 21486-032 DIV5

CURRENT APPLICATION NUMBER: US/09/903,248

PRIOR FILING DATE: 2001-07-11

PRIOR APPLICATION NUMBER: 09/436,184

PRIOR FILING DATE: 1999-11-08

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 758

TYPE: PRT

ORGANISM: Homo sapiens

US-09-903-248-2

Query Match	100.0%	Score 4022;	DB 9;	Length 758;
Best Local Similarity	100.0%;	Pred. No. 1.5e-282;		
Matches 758;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAQRKNAKSSGNSSSSGSGSTSAGSSSPGARRETKHGCHKNGRKGGLSGTSFTFMV	60	
Db	1	MAQRKNAKSSGNSSSSGSGSTSAGSSSPGARRETKHGCHKNGRKGGLSGTSFTFMV	60	
QY	61	TALLGWTSAVAVWFDLVDYBEVLGKLTGYDADGDGDFDVKAKYLLGKERSTSEPAVP	120	
Db	61	TALLGWTSAVAVWFDLVDYBEVLGKLTGYDADGDGDFDVKAKYLLGKERSTSEPAVP	120	
QY	121	PEAEPTPEPEQVVEAEFPONTEDEAKEOIOSLIHEMVHAHEHVEGEDLQOEDGPTGEPO	180	
Db	121	PEAEPTPEPEQVVEAEFPONTEDEAKEOIOSLIHEMVHAHEHVEGEDLQOEDGPTGEPO	180	

```

QY 181 QEDDEFLMATVDVDRFETLEPEVSHHEETESHVHEETVSQDCNQDMEMMSEQENDSSSE 240
Db 181 QEDDEFLMATVDVDRFETLEPEVSHHEETESHVHEETVSQDCNQDMEMMSEQENDSSSE 240
QY 241 PVVEDERLHHDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS I 300
Db 241 PVVEDERLHHDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS I 300
QY 301 FPVEEQOEVPPEETNRKTTDDPEQAKAVKKKKPKLNFDKTIKALDAAEKLRKRGKIEEA 360
Db 301 FPVEEQOEVPPEETNRKTTDDPEQAKAVKKKKPKLNFDKTIKALDAAEKLRKRGKIEEA 360
QY 361 VNAFKELVRKYPOSPPARARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVADLLK 420
Db 361 VNAFKELVRKYPOSPPARARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVADLLK 420
QY 421 LSLKRSDRQOFLGHMRGSLTLQRLVOLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
Db 421 LSLKRSDRQOFLGHMRGSLTLQRLVOLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
QY 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGDFYFHLGDAMQRVGN 540
Db 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGDFYFHLGDAMQRVGN 540
QY 541 KEAYKWYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKWYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTTLLEKPEBTTC 660
Db 601 GLAVMDKAKGFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTTLLEKPEBTTC 660
QY 661 RRGQIKYSIMHGTWHPHTGPTNCRMLGLVIPKEGCKIRCANETRTWEEGKYLIFD 720
Db 661 RRGQIKYSIMHGTWHPHTGPTNCRMLGLVIPKEGCKIRCANETRTWEEGKYLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758
Db 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758

```

```

RESULT 2
US-09-859-604-2
; Sequence 2, Application US/09859604
; Patent No. US20020110559A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M
; APPLICANT: Deutch, Alan H
; APPLICANT: Chanbari, Hossein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; PRIOR FILING DATE: 2001-05-17
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-859-604-2

```

```

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAQRKNAKSSGNSGSSGSGSGSTSAGSGSGPGARRETKHGHKNGRKGLSGTSFPTWVMY 60
Db 1 MAQRKNAKSSGNSGSSGSGSGSTSAGSGSGPGARRETKHGHKNGRKGLSGTSFPTWVMY 60

```

```

QY 61 IALLGWTSAVAVWFDLVYEEVLGKLGITYDADGDFDVDDAKVLLGLKERSTSEPAVP 120
Db 61 IALLGWTSAVAVWFDLVYEEVLGKLGITYDADGDFDVDDAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEPOINIEDEAKEQIQSLHENVHAHEVEGEDLQOEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVPVEAEPOINIEDEAKEQIQSLHENVHAHEVEGEDLQOEDGPTGEPQ 180
QY 181 QEDDEFLMATVDVDRFETLEPEVSHHEETESHVHEETVSQDCNQDMEMMSEQENDSSSE 240
Db 181 QEDDEFLMATVDVDRFETLEPEVSHHEETESHVHEETVSQDCNQDMEMMSEQENDSSSE 240
QY 241 PVVEDERLHHDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS I 300
Db 241 PVVEDERLHHDTDVTVQVVEEQAVVEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS I 300
QY 301 FPVEEQOEVPPEETNRKTTDDPEQAKAVKKKKPKLNFDKTIKALDAAEKLRKRGKIEEA 360
Db 301 FPVEEQOEVPPEETNRKTTDDPEQAKAVKKKKPKLNFDKTIKALDAAEKLRKRGKIEEA 360
QY 361 VNAFKELVRKYPOSPPARARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVADLLK 420
Db 361 VNAFKELVRKYPOSPPARARYGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVADLLK 420
QY 421 LSLKRSDRQOFLGHMRGSLTLQRLVOLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
Db 421 LSLKRSDRQOFLGHMRGSLTLQRLVOLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
QY 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGDFYFHLGDAMQRVGN 540
Db 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDGDFYFHLGDAMQRVGN 540
QY 541 KEAYKWYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKWYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTTLLEKPEBTTC 660
Db 601 GLAVMDKAKGFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTTLLEKPEBTTC 660
QY 661 RRGQIKYSIMHGTWHPHTGPTNCRMLGLVIPKEGCKIRCANETRTWEEGKYLIFD 720
Db 661 RRGQIKYSIMHGTWHPHTGPTNCRMLGLVIPKEGCKIRCANETRTWEEGKYLIFD 720
QY 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758
Db 721 DSFEHEVWQDASSFRLLIFIVDVWHPBELTPQORSLPAI 758

```

```

RESULT 3
US-09-903-063-2
; Sequence 2, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; PRIOR FILING DATE: 2001-10-11
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-063-2

```

```

Query Match 100.0%; Score 4022; DB 9; Length 758;

```

```
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MAQRKNAKSSGSSSSGSGSGSTSAGSSSPGARRETKHGHNKGRKGGLSGTSFTWEMV 60
Db      1 MAQRKNAKSSGSSSSGSGSGSTSAGSSSPGARRETKHGHNKGRKGGLSGTSFTWEMV 60
QY      61 IALLGVMTSAVWVFDLVDYEEVLGKLGITYDADGDGDFVDPAKVLGLKERSTSEPAVP 120
Db      61 IALLGVMTSAVWVFDLVDYEEVLGKLGITYDADGDGDFVDPAKVLGLKERSTSEPAVP 120
QY      121 PEEAEPHTEPEEQVPVEAEPQNIIDEAKQIQSLHEMVHAHVEGEDLQOEDGPTGEPQ 180
Db      121 PEEAEPHTEPEEQVPVEAEPQNIIDEAKQIQSLHEMVHAHVEGEDLQOEDGPTGEPQ 180
QY      181 QEDDEFMATVDVDRFETLEPEVSHETESHVHEETVSQDCNQDMEEEMSEQENDPSSE 240
Db      181 QEDDEFMATVDVDRFETLEPEVSHETESHVHEETVSQDCNQDMEEEMSEQENDPSSE 240
QY      241 PVVEDERLHDDTDVTTYQVVEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
Db      241 PVVEDERLHDDTDVTTYQVVEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
QY      301 FPVEEQOEVPEETNRKTDDEQAKAVKKKKPKLNFDPKTIKAEFDAEKLRRKGKIEEA 360
Db      301 FPVEEQOEVPEETNRKTDDEQAKAVKKKKPKLNFDPKTIKAEFDAEKLRRKGKIEEA 360
QY      361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
Db      361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
QY      421 LSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVCYLLIGDNDNAKYYEEV 480
Db      421 LSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVCYLLIGDNDNAKYYEEV 480
QY      481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
Db      481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
QY      541 KEAYKMYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
Db      541 KEAYKMYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY      601 GLAVMDKAKGLFLPEDENLREKQDWSQFTLWQGRRNENACKGAPKCTTLEKPEETGCG 660
Db      601 GLAVMDKAKGLFLPEDENLREKQDWSQFTLWQGRRNENACKGAPKCTTLEKPEETGCG 660
QY      661 RRGQIKYSIMHPGTHWPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720
Db      661 RRGQIKYSIMHPGTHWPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720
QY      721 DSFEHEVWQDASSFRLIFIVDVWHPBELTPQQRSLPAI 758
Db      721 DSFEHEVWQDASSFRLIFIVDVWHPBELTPQQRSLPAI 758

RESULT 4
US-09-903-216-2
; Sequence 2, Application US/09903216
; Patent No. US20020114811A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV2
; CURRENT APPLICATION NUMBER: US/09/903,216
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-216-2

Query Match      100.0%; Score 4022; DB 9; length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MAQRKNAKSSGSSSSGSGSGSTSAGSSSPGARRETKHGHNKGRKGGLSGTSFTWEMV 60
Db      1 MAQRKNAKSSGSSSSGSGSGSTSAGSSSPGARRETKHGHNKGRKGGLSGTSFTWEMV 60
QY      61 IALLGVMTSAVWVFDLVDYEEVLGKLGITYDADGDGDFVDPAKVLGLKERSTSEPAVP 120
Db      61 IALLGVMTSAVWVFDLVDYEEVLGKLGITYDADGDGDFVDPAKVLGLKERSTSEPAVP 120
QY      121 PEEAEPHTEPEEQVPVEAEPQNIIDEAKQIQSLHEMVHAHVEGEDLQOEDGPTGEPQ 180
Db      121 PEEAEPHTEPEEQVPVEAEPQNIIDEAKQIQSLHEMVHAHVEGEDLQOEDGPTGEPQ 180
QY      181 QEDDEFMATVDVDRFETLEPEVSHETESHVHEETVSQDCNQDMEEEMSEQENDPSSE 240
Db      181 QEDDEFMATVDVDRFETLEPEVSHETESHVHEETVSQDCNQDMEEEMSEQENDPSSE 240
QY      241 PVVEDERLHDDTDVTTYQVVEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
Db      241 PVVEDERLHDDTDVTTYQVVEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
QY      301 FPVEEQOEVPEETNRKTDDEQAKAVKKKKPKLNFDPKTIKAEFDAEKLRRKGKIEEA 360
Db      301 FPVEEQOEVPEETNRKTDDEQAKAVKKKKPKLNFDPKTIKAEFDAEKLRRKGKIEEA 360
QY      361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
Db      361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
QY      421 LSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVCYLLIGDNDNAKYYEEV 480
Db      421 LSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVCYLLIGDNDNAKYYEEV 480
QY      481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
Db      481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
QY      541 KEAYKMYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
Db      541 KEAYKMYELGHRGHFASVWQRSLYNVNGLKAQPMWTPKETGYTELKSLERNWKLIRDE 600
QY      601 GLAVMDKAKGLFLPEDENLREKQDWSQFTLWQGRRNENACKGAPKCTTLEKPEETGCG 660
Db      601 GLAVMDKAKGLFLPEDENLREKQDWSQFTLWQGRRNENACKGAPKCTTLEKPEETGCG 660
QY      661 RRGQIKYSIMHPGTHWPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720
Db      661 RRGQIKYSIMHPGTHWPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720
QY      721 DSFEHEVWQDASSFRLIFIVDVWHPBELTPQQRSLPAI 758
Db      721 DSFEHEVWQDASSFRLIFIVDVWHPBELTPQQRSLPAI 758

RESULT 5
US-09-903-199-2
; Sequence 2, Application US/09903199
; Patent No. US20020122802A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
```

FILE REFERENCE: 21486-032 DIV4
CURRENT APPLICATION NUMBER: US/09/903,199
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/436,184
PRIOR FILING DATE: 1999-11-08
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 758
TYPE: PRT
ORGANISM: Homo sapiens
US-09-903-199-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGSSSSSGSGSSTAGSSSPGARRETGHGKNGRKGGLSGTSFPTWFW 60
Db 1 MAQRKNAKSSGSSSSSGSGSSTAGSSSPGARRETGHGKNGRKGGLSGTSFPTWFW 60
QY 61 IALLGVTSAVAVWFVFDLVYEEVLGKLGIYDADGDGDFVDADAKVLLGLKERSTSEPAVP 120
Db 61 IALLGVTSAVAVWFVFDLVYEEVLGKLGIYDADGDGDFVDADAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPEAEAPQNIIDEAKEQIQSLHEMVHAHEVGEDLQOEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVPEAEAPQNIIDEAKEQIQSLHEMVHAHEVGEDLQOEDGPTGEPQ 180
QY 181 QEDDEFMATDVDDREFETLEPEVSHETESHVHEETVSQDCNQMEEEMSEQENPDSSSE 240
Db 181 QEDDEFMATDVDDREFETLEPEVSHETESHVHEETVSQDCNQMEEEMSEQENPDSSSE 240
QY 241 PVVEDERLHDDTDVTTYQVYEEQAVVEPLENEGIBITEVTAPREDNPVEDSQVIVEEVS 300
Db 241 PVVEDERLHDDTDVTTYQVYEEQAVVEPLENEGIBITEVTAPREDNPVEDSQVIVEEVS 300
QY 301 FPVEEQEQVPPETNRKTDDEQAKAVKKKKPKLNFDPKTIKALDAEKLRRKGKIEEA 360
Db 301 FPVEEQEQVPPETNRKTDDEQAKAVKKKKPKLNFDPKTIKALDAEKLRRKGKIEEA 360
QY 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
Db 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
QY 421 LSLKRSDRQOQFLGMRGSLTLQRLVQLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
Db 421 LSLKRSDRQOQFLGMRGSLTLQRLVQLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
QY 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQVRGN 540
QY 541 KEAYKMYELGHRGHFASVWQSLYVNGIKAKQPMWTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKMYELGHRGHFASVWQSLYVNGIKAKQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQGRRNENACKGAPKCTCTLLEKPEPTTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQGRRNENACKGAPKCTCTLLEKPEPTTGC 660
QY 661 RRGQIKYSIMHPGTHWPHGTPTNCRMLHGLVIPKEGCKIRCANETRWEKGVLIFD 720
Db 661 RRGQIKYSIMHPGTHWPHGTPTNCRMLHGLVIPKEGCKIRCANETRWEKGVLIFD 720
QY 721 DSFEHEVWQDASSFRLIFIVDVWHPDELTPQQRSLPAI 758
Db 721 DSFEHEVWQDASSFRLIFIVDVWHPDELTPQQRSLPAI 758

RESULT 6
US-09-903-023-2
Sequence 2, Application US/09903023

Patent No. US20020146421A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carleon, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: 21486-032 DIV1
CURRENT APPLICATION NUMBER: US/09/903,023
CURRENT FILING DATE: 2001-10-11
PRIOR APPLICATION NUMBER: 09/436,184
PRIOR FILING DATE: 1999-11-08
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 758
TYPE: PRT
ORGANISM: Homo sapiens
US-09-903-023-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGSSSSSGSGSSTAGSSSPGARRETGHGKNGRKGGLSGTSFPTWFW 60
Db 1 MAQRKNAKSSGSSSSSGSGSSTAGSSSPGARRETGHGKNGRKGGLSGTSFPTWFW 60
QY 61 IALLGVTSAVAVWFVFDLVYEEVLGKLGIYDADGDGDFVDADAKVLLGLKERSTSEPAVP 120
Db 61 IALLGVTSAVAVWFVFDLVYEEVLGKLGIYDADGDGDFVDADAKVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPEAEAPQNIIDEAKEQIQSLHEMVHAHEVGEDLQOEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVPEAEAPQNIIDEAKEQIQSLHEMVHAHEVGEDLQOEDGPTGEPQ 180
QY 181 QEDDEFMATDVDDREFETLEPEVSHETESHVHEETVSQDCNQMEEEMSEQENPDSSSE 240
Db 181 QEDDEFMATDVDDREFETLEPEVSHETESHVHEETVSQDCNQMEEEMSEQENPDSSSE 240
QY 241 PVVEDERLHDDTDVTTYQVYEEQAVVEPLENEGIBITEVTAPREDNPVEDSQVIVEEVS 300
Db 241 PVVEDERLHDDTDVTTYQVYEEQAVVEPLENEGIBITEVTAPREDNPVEDSQVIVEEVS 300
QY 301 FPVEEQEQVPPETNRKTDDEQAKAVKKKKPKLNFDPKTIKALDAEKLRRKGKIEEA 360
Db 301 FPVEEQEQVPPETNRKTDDEQAKAVKKKKPKLNFDPKTIKALDAEKLRRKGKIEEA 360
QY 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
Db 361 VNAFKELVRKYPQSPRARYGKAOCEDDLAEKRRSNEVLRGAIETTYQEVASLPDVPADLLK 420
QY 421 LSLKRSDRQOQFLGMRGSLTLQRLVQLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
Db 421 LSLKRSDRQOQFLGMRGSLTLQRLVQLFPNDTSLKNDLGVGYLIGDNDNAKKYEEV 480
QY 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQVRGN 540
QY 541 KEAYKMYELGHRGHFASVWQSLYVNGIKAKQPMWTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKMYELGHRGHFASVWQSLYVNGIKAKQPMWTPKETGYTELKSLERNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQGRRNENACKGAPKCTCTLLEKPEPTTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQGRRNENACKGAPKCTCTLLEKPEPTTGC 660
QY 661 RRGQIKYSIMHPGTHWPHGTPTNCRMLHGLVIPKEGCKIRCANETRWEKGVLIFD 720
Db 661 RRGQIKYSIMHPGTHWPHGTPTNCRMLHGLVIPKEGCKIRCANETRWEKGVLIFD 720
QY 721 DSFEHEVWQDASSFRLIFIVDVWHPDELTPQQRSLPAI 758
Db 721 DSFEHEVWQDASSFRLIFIVDVWHPDELTPQQRSLPAI 758

Db 721 DSFEHEWQDASSFRLIFVDVWHPDLTPQORSLPAI 758

RESULT 7

US-09-436-184-2
; Sequence 2, Application US/09436184
; Publication No. US20030031670A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: R.I. Hosp. - Malignant Neoplasms
; CURRENT APPLICATION NUMBER: US/09/436,184
; CURRENT FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-436-184-2

Query Match 100.0%; Score 4022; DB 10; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRNNAKSSGNSSSGSGSGSTASGSSSPGARRETKHGKNGRKGISGTSFPTWFMV 60
Db 1 MAQRNNAKSSGNSSSGSGSGSTASGSSSPGARRETKHGKNGRKGISGTSFPTWFMV 60
QY 61 IALIGWTSVAVVWFVLDVYEEVLGKLGITYDADGDGDFDVAKVLGLKERSTSEPAVP 120
Db 61 IALIGWTSVAVVWFVLDVYEEVLGKLGITYDADGDGDFDVAKVLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPEAEQONIEDAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGEPO 180
Db 121 PEEAEPHTEPEEQVPEAEQONIEDAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGEPO 180
QY 181 QEDDEFIMATVDVDRFETLEPEVSHETESHYHEETVSQDCNODMEEMSEQENPDSS 240
Db 181 QEDDEFIMATVDVDRFETLEPEVSHETESHYHEETVSQDCNODMEEMSEQENPDSS 240
QY 241 PVVEDERLHMDTDVTVQVYEEQAVYEPLENEGIEITEVTAPREDNPVEDSQVIVEEVS 300
Db 241 PVVEDERLHMDTDVTVQVYEEQAVYEPLENEGIEITEVTAPREDNPVEDSQVIVEEVS 300
QY 301 FVPEEQEVPPETNRKTDDEQAKAVKKPKLNFKDTIKAEILDAEKLRRKGIIEEA 360
Db 301 FVPEEQEVPPETNRKTDDEQAKAVKKPKLNFKDTIKAEILDAEKLRRKGIIEEA 360
QY 361 VNAFELVRKYPQSPRARYGAQCEDDLAEKRRSNEVLGAIEITYQEVASLPDVPADLTK 420
Db 361 VNAFELVRKYPQSPRARYGAQCEDDLAEKRRSNEVLGAIEITYQEVASLPDVPADLTK 420
QY 421 LSLKRSDRQOFLGHMRGSLITLQRLVOLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
Db 421 LSLKRSDRQOFLGHMRGSLITLQRLVOLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
QY 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPLYKEGIESGDPGTDGFRFYHLGDAMQVRGN 540
Db 481 LSVTPNDGFAKVHYGFILKAQNKIAESIPLYKEGIESGDPGTDGFRFYHLGDAMQVRGN 540
QY 541 KEAYKMYELGHKRGHFASVQWQSLYNVNGLKAQPMWTPKETGYTELVSLENNWKLIRDE 600
Db 541 KEAYKMYELGHKRGHFASVQWQSLYNVNGLKAQPMWTPKETGYTELVSLENNWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENIREKGDWSQFTLWQGRNENACKGAPKTCITLLEKFPETTG 660
Db 601 GLAVMDKAKGLFLPEDENIREKGDWSQFTLWQGRNENACKGAPKTCITLLEKFPETTG 660

QY 661 RRGQIKYSIMHPGTHVPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720
Db 661 RRGQIKYSIMHPGTHVPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEGKVLIFD 720

QY 721 DSFEHEWQDASSFRLIFVDVWHPDLTPQORSLPAI 758

RESULT 8

US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaltanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1806

Query Match 32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVIALIGWTSVAVVWFVLDVYEEVLGKLGITYDADGDGDFDVAKVLGLKERSTSEPA 118
Db 1 MVIALIGWTSVAVVWFVLDVYEEVLGKLGITYDADGDGDFDVAKVLGLKERSTSEPA 60
QY 119 VPPEAEPHTEPEEQVPEAEQONIEDAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGE 178
Db 61 VPPEAEPHTEPEEQVPEAEQONIEDAKEQIOSLHEMVHAHEVEGEDLOQEDGPTGE 120
QY 179 PQEEDDEFIMATVDVDRFETLEPEVSHETESHYHEETVSQDCNODMEEMSEQENPD 238
Db 121 PQEEDDEFIMATVDVDRFETLEPEVSHETESHYHEETVSQDCNODMEEMSEQENPD 180
QY 239 SEPVEDERLHMDTDVTVQVYEEQAVYEPLENEGIEITEVTAPREDNPVEDSQVIVEE 298
Db 181 SEPVEDERLHMDTDVTVQVYEEQAVYEPLENEGIEITEVTAPREDNPVEDSQVIVEE 240
QY 299 SIFPVEEQEVPPET 313
Db 241 SIFPVEEQEVPPET 255

RESULT 9

US-09-902-941-1806
; Sequence 1806, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita


```
Db      241 SIFPVEEQEVPDPDT 255

RESULT 12
US-10-017-754-1806
; Sequence 1806, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnierakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVKVLLGLKERSTSEPA 118
      |||
Db      1 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVKVLLGLKERSTSEPA 60

QY      119 VPPEAEHPTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVHAHEVGEEDLQOEDGPTGE 178
      |||
Db      61 VPPEAEHPTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVHAHEVGEEDLQOEDGPTGE 120

QY      179 PQOEDDEFIMATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEEMNSEQENPDS 238
      |||
Db      121 PQOEDDEFIMATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEEMNSEQENPDS 180

QY      239 SEPVEDERLHDDTDVTQVVEEQAVYEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 298
      |||
Db      181 SEPVEDERLHDDTDVTQVVEEQAVYEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 240

QY      299 SIFPVEEQEVPDPDT 313
      |||
Db      241 SIFPVEEQEVPDPDT 255

RESULT 13
US-10-113-872-1806
; Sequence 1806, Application US/10113872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C19
; CURRENT APPLICATION NUMBER: US/10/113,872
```

```
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-113-872-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVKVLLGLKERSTSEPA 118
      |||
Db      1 MVALLGWTSVAVWVFDLVYEEVLGKLGITYDADGDFDVKVLLGLKERSTSEPA 60

QY      119 VPPEAEHPTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVHAHEVGEEDLQOEDGPTGE 178
      |||
Db      61 VPPEAEHPTEPEEQVPEAEPPQNIIDEAKEQIQSLHEMVHAHEVGEEDLQOEDGPTGE 120

QY      179 PQOEDDEFIMATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEEMNSEQENPDS 238
      |||
Db      121 PQOEDDEFIMATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEEMNSEQENPDS 180

QY      239 SEPVEDERLHDDTDVTQVVEEQAVYEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 298
      |||
Db      181 SEPVEDERLHDDTDVTQVVEEQAVYEPLNEGIEITEVTAPEDNPVEDSQVIVEEV 240

QY      299 SIFPVEEQEVPDPDT 313
      |||
Db      241 SIFPVEEQEVPDPDT 255

RESULT 14
US-10-156-761-8153
; Sequence 8153, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Query Match      6.8%; Score 275; DB 14; Length 250;
Best Local Similarity 33.3%; Pred. No. 9.3e-12;
Matches 65; Conservative 31; Mismatches 91; Indels 8; Gaps 4;

QY      569 GLKAQPPWTPKKTGYTEL---YKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDW 625
      |||
Db      48 GLSPTPWHPD--YAYAEAPVYVHELEASHQAIXELKVAMSARREAFSDYEHVLTRODNW 105

QY      626 SQFTLMQGRNENACKGAPKCTLLLEKFPETG--CRGQIKYSIMHPGTHVPHPTGPT 683
      |||
Db      106 QALYLFREGGITBESATVPPIAYQVLKDVAVDTGKICPLLCHFTLLPGAVIEPHCDLW 165
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:54:54 ; Search time 31.7034 Seconds
(without alignments)
2299.852 Million cell updates/sec

Title: US-09-903-199-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSSGSGS.....IVDVWHELTPOQRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	3966.5	98.6	757	2	I38423 aspartyl beta-hydr
2	3172	78.9	754	1	BABOH peptide-aspartate
3	890	22.1	270	2	UC7792 cardiac junctate-1
4	798	19.8	872	2	T18861 probable peptide-a
5	263.5	6.6	186	2	T47148 hypothetical prote
6	216.5	5.4	1110	2	I51116 NF-180 - sea lamp
7	199	4.9	1271	2	A45555 glutamate rich pro
8	197.5	4.9	312	2	H83527 hypothetical prote
9	191.5	4.8	1881	2	H95076 zinc metalloprotei
10	189.5	4.7	1616	2	G64242 cytodherence-acces
11	188.5	4.7	792	2	T42963 hypothetical prote
12	187	4.6	706	2	A45990 junctional sarcopl
13	183	4.5	763	2	T08929 hypothetical prote
14	181	4.5	771	1	A33430 h-caldesmon - chic
15	179.5	4.5	411	2	S47436 flagellar antigen
16	178.5	4.4	729	2	S68191 triadin - human
17	177	4.4	301	2	A82601 aspartyl/asparagin
18	177	4.4	845	2	A45669 neurofilament trip
19	176	4.4	1094	2	S49313 protein kinase - s
20	174.5	4.3	1948	2	S00485 gene 11-1 protein
21	173.5	4.3	1871	2	D96796 probable heat shoc
22	172.5	4.3	1240	2	S52734 hypothetical prote
23	172	4.3	465	2	S46759 hypothetical prote
24	172	4.3	630	2	S29796 hypothetical prote
25	171.5	4.3	607	2	S27776 80K protein (allele
26	171	4.3	1804	2	T34518 nestin - golden ha
27	169.5	4.2	592	2	B48315 lamin B2 - mouse
28	169.5	4.2	1233	2	S56271 hypothetical prote
29	168.5	4.2	695	2	T40168 hypothetical prote

30	168.5	4.2	913	2	T52485 neurofilament prot
31	168	4.2	302	2	AF1021 probable membrane-
32	168	4.2	1192	2	A71623 probable secreted
33	167.5	4.2	1876	2	B97944 zinc metalloprotei
34	166.5	4.1	1624	2	T25592 hypothetical prote
35	165	4.1	651	2	S18874 nucleolin - Africa
36	164.5	4.1	7962	2	I38346 elastic titin - hu
37	164	4.1	1957	2	T38077 hypothetical coile
38	163.5	4.1	1132	2	T43483 translation initia
39	163	4.1	411	2	S48647 peptidylprolyl iso
40	163	4.1	501	2	C71948 hypothetical prote
41	163	4.1	1877	2	T21861 hypothetical prote
42	163	4.1	1999	1	S21801 myosin heavy chain
43	162.5	4.0	299	2	B83081 hypothetical prote
44	162.5	4.0	793	1	JH0628 caldesmon - human
45	162.5	4.0	992	2	T46337 hypothetical prote

ALIGNMENTS

RESULT 1
I38423 aspartyl beta-hydroxylase - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000
C:Accession: I38423
R:Koricho, F.; Gieffers, C.; Frey, J.
Gene 150, 395-399, 1994
A:Title: Cloning and characterization of the human gene encoding aspartyl beta-hydroxyl.
A:Reference number: I38423; MUID:95121937; PMID:7821814
A:Accession: I38423
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-757 <RES>
A:Cross-references: EMBL:U03109; NID:G458031; PIDD:AAA82108.1; PID:G458032
C:Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology
F:54-75/Domain: transmembrane #status predicted <TRM>
F:341-374/Domain: tetratricopeptide repeat homology <TRR>

Query Match	Best Local Similarity	Score	DB 2;	Length	757;
Matches	752;	Conservative	2;	Mismatches	3;
				Indels	1;
				Gaps	1;
QY	1	MAQRKNAKSSGNSSSSSGSGS	TAAGSPPGARRRTHGHHKNGRKGGLSGTSPTFMV	60	
DB	1	MAQRKNAKSSGNSSSSSGSGS	TASGSSPGARRRTHGHHKNGRKGGLSGTSPTFMV	60	
QY	61	IALLGWTVAVVMFDLVYBEVLGKLGIYDADGGDFVDVDAKVLGLKERSTSEPAVP	120		
DB	61	IALLGWTVAVVMFDLVYBEVLGKLGIYDADGGDFVDVDAKVLGLKERSTSEPAVP	120		
QY	121	PEEAPHTPEPEEQVPVEAEPONIDEAKEQIQSLHEMVAHEVEGEDLQEDGPTGEPQ	180		
DB	121	PEEAPHTPEPEEQVPVEAEPONIDEAKEQIQSLHEMVAHEVEGEDLQEDGPTGEPQ	180		
QY	181	QEDDEFMATDVDRFETLEPEVSHETESHYHVEETVSQDCNQDMEEHMSQENPDSSSE	240		
DB	181	QEDDEFMATDVDRFETLEPEVSHETESHYHVEETVSQDCNQDMEEHMSQENPDSSSE	240		
QY	241	PVVEDERLHHTDDVTYQVVEEQAVYERPLENEGIEITEVTAPBEDNPVEDSQVIEVSI	300		
DB	241	PVVEDERLHHTDDVTYQVVEEQAVYERPLENEGIEITEVTAPBEDNPVEDSQVIEVSI	300		
QY	301	FPVEEQEQVPPETNRTKTDDEPEQAKVKKKKPKLNFQDKTIKAEILDAEKLKRKRIEBA	360		
DB	301	FPVEEQEQVPPETNRTKTDDEPEQAKVKKKKPKLNFQDKTIKAEILDAEKLKRKRIEBA	360		
QY	361	VNAFKELVRKYPQSPRARYGKQCEDDLAEKRRSNEVLGAIETTYQEVASLPDVPADLLK	420		
DB	361	VNAFKELVRKYPQSPRARYGKQCEDDLAEKRRSNEVLGAIETTYQEVASLPDVPADLLK	420		
QY	421	LSLKRSDRQQLGHMRGSLTLQLRVQLFPNDTSLKNDLGVGYLLIGDNDNAKKYIEEV	480		

Db 421 LSLKRRSDRQQLGHNMGSLTLQLVQLFENDTSLKNDLGVGYLLIGDNDNAKKVBEV 480
QY 481 LSVTPNDGFAKVHYGFLKKAQNKIAESIPYLKEGIESGDPGTDDGRFFYHLDAMQRYGN 540
Db 481 LSVTPNDGFAKVHYGFLKKAQNKIAESIPYLKEGIESGDPGTDDGRFFYHLDAMQRYGN 540
QY 541 KEAYKMYELGHRGHFASVWQSLYNVNGLKAPWMTPKETGYTELKSLERNWKLIRDE 600
Db 541 KEAYKMYELGHRGHFASVWQSLYNVNGLKAP-CGPKETGYTQLVKSLEARNWKLIRDE 599
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRNENACKGAPKTCTLLEKFPETTG 660
Db 600 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRNENACKGAPKTCTLLEKFPETTG 659
QY 661 RRGQIKYSIMHPTGTHVPHPTGPTNCRRLRMHGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
Db 660 RRGQIKYSIMHPTGTHVPHPTGPTNCRRLRMHGLVIPKEGCKIRCANETRTWEEGKVLIFD 719
QY 721 DSFEHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 758
Db 720 DSFEHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 757

RESULT 2

BABOH
peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine
N:Alternate names: aspartyl (asparaginyl) beta-hydroxylase
C:Species: Bos primigenius taurus (cattle)
C>Date: 31-Dec-1993 #sequence revision 10-Feb-1995 #text_change 11-Jun-1999
C:Accession: A42969; A39470; B39470; C39470; S27948
R:Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.B.; Dixon, R.A.; Elliston, K.O.; Stern, J. Biol. Chem. 267, 14322-14327, 1992
A:Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.
A:Reference number: A42969; MUID:92332546; PMID:1378441
A:Accession: A42969
A:Molecule type: mRNA
A:Residues: 1-754 <JIA>
A:Cross-references: EMBL:M91213; NID:g162693; PIDN:AAA03563.1; PID:g162694
A:Experimental source: Brain
A>Note: sequence extracted from NCBI backbone (NCBIP:108534)
R:Wang, Q.; Vandusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A. J. Biol. Chem. 266, 14004-14010, 1991
A:Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.
A:Reference number: A39470; MUID:91310689; PMID:1856229
A:Accession: A39470
A:Molecule type: protein
A:Residues: 289-328 <WAN>
A:Accession: B39470
A:Molecule type: protein
A:Residues: 615, 'X', 617-630, 'XX', 633-634, 'X', 636, 'XX', 639-641 <WA2>
A:Accession: C39470
A:Molecule type: protein
A:Residues: 311-347, 'X', 349, 'X', 351-373, 'X', 375-379, 'X', 381-382 <WA3>
C:Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the C:Superfamily: Aspartic acid and asparagine residues in the EGF homology domain of certain F C:Keywords: glycoprotein; oxidoreductase; transmembrane protein
F:2-56/Domain: intracellular #status predicted <INC>
F:57-78/Domain: transmembrane #status predicted <TRM>
F:289-754/Product: peptide-aspartate beta-dioxygenase, 56K form #status predicted <56K>
F:311-754/Product: peptide-aspartate beta-dioxygenase, 52K form #status predicted <52K>
F:337-370/Domain: tetratricopeptide repeat homology <TT1>
F:371-404/Domain: tetratricopeptide repeat homology <TT2>
F:13, 96, 466, 702/Binding site: carbonylate (Aen) (covalent) #status predicted

Query Match

Best Local Similarity 78.9%; Score 3172; DB 1; Length 754;
Matches 612; Conservative 45; Mismatches 77; Indels 44; Gaps 6;

QY 1 MAQRNKAASSG--NSSSSGSGSGSTASGSSPGARRRTHGCHKRKGKGLSGTSFTTW 57
Db 1 MAQRNKAASSG--NSSSSGSGSGSGSTASGSSPGARRRTHGCHKRKGKGLSGTSFTTW 60

QY 58 FMVIALLGWTSVAVWFDLVDYBEVL-----GKLGIDADGDGDEVD 102
Db 61 FMVIALLGWTSVAVWFDLVDYBEVLAKAKDFRNYLSEVLQKLGIDADGDGDEVD 120
QY 103 AKVLLGKERSTSEPAVPPEEADPHTEPEQVPVEAEPOINDEAKEQIOSLHEWHA 162
Db 121 AKVLLGKERPAKPTVPPEADWYPLLEDVLESFGRONIDEVYEQVS-LDETVYSE 179
QY 163 HVEGEDLQOEDGPTGEPOQDEDEFLMATDVDDRFETLEPEVSHETEHSYHVEETVSGDC 222
Db 180 --PGENLPQEPGPAEELQPDHVFVGSDDADDRYEPMTGAVHEETEDSYHIEETASPAY 237
QY 223 NQDMEEMSEQENPDSSPEVVED--ERLHDTDDVTYQVYEQAVYEPLENEGIEETVT 280
Db 238 SQDMEEMMYEQENPDSSPEVVDAAERTYQETDDVTYRDYDEQ----- 280
QY 281 APPEDNPVEDSQVIVEEVSIFPVEEQEVPEPNRKTDDPEQAKVKKKPKLNFDPKT 340
Db 281 ---DHAVDNSNTLLEPHMPPAEQEVPPETNKKADEPGKKGVKKKPKLNFDPKT 336
QY 341 IKAELDAAEKLRKRGKIEAVNAFKELVKYPQSPRAYGKAQCEDDLAEKRRSNEVL 400
Db 337 IKAELDAAEKLRKRGKIEAVNAFKELVKYPQSPRAYGKAQCEDDLAEKRRSNEVL 396
QY 401 AIETYEAVASLPDVPADLKLTLKRRSDRQQLGHNMGSLTLQLVQLFNDTSLKNDL 460
Db 397 AIETYEAVASLPDAPTDLVKLSLKRSDRQQLGHNMGSLTLQLVQLFDDTALKNDL 456
QY 461 GVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFLKKAQNKIAESIPYLKEGIESGDP 520
Db 457 GVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFLKKAQNKIAESIPYLKEGIESGDP 516
QY 521 GTDDGRFFYHLDAMQORVGNKEAYKMYELGHRGHFASVWQSLYNVNGLKAPWMTPK 580
Db 517 GTDDGRFFYHLDAMQORVGNKEAYKMYELGHRGHFASVWQSLYNVNGLKAPWMTPK 576
QY 581 TGYTELKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQOGRNENA 640
Db 577 TGYTELKSLERNWKLIRDEGLAAMDRTGHLFLPEDENLREKGDWSQFTLMQOGRNENA 636
QY 641 CKGAPKTCTLLEKFPETTGCRGQIKYSIMHPTGTHVPHPTGPTNCRRLRMHGLVIPKEGC 700
Db 637 CKGAPKTCTLLEKFPETTGCRGQIKYSIMHPTGTHVPHPTGPTNCRRLRMHGLVIPKEGC 696
QY 701 KIRCANETRTWEEGKVLIFDDSFHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 758
Db 697 KIRCANETRTWEEGKVLIFDDSFHEVWQDASSFRLLIFVDVWHPHPELTPOQRSLPAI 754

RESULT 3

JC7792
cardiac junctate-1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 03-May-2002
C:Accession: JC7792
R:Hong, C.S.; Kwak, Y.G.; Ji, J.H.; Chae, S.W.; Kim, D.H. Biochem. Biophys. Res. Commun. 289, 882-887, 2001
A:Title: Molecular cloning and characterization of mouse cardiac junctate isoforms.
A:Reference number: JC7792
A:Contents: Heart
A:Accession: JC7792
A:Molecule type: mRNA
A:Residues: 1-270 <HON>
A:Cross-references: GB:AF302653
C:Comment: This protein, a Ca2+ binding protein. plays a role both in contractile and c
C:Keywords: cardiac muscle; heart

Query Match

Best Local Similarity 22.1%; Score 890; DB 2; Length 270;
Matches 179; Conservative 30; Mismatches 57; Indels 14; Gaps 5;

QY 34 RETKHGHRKGRKGKGLSGTSFTTWFMVIALLGWTSVAVWFDLVDYBEVLGKLGIDAD 93

Db	5	KEAKHGKHNGRRGGISGSGSFFTWFWYIALLGWTSVAVWFMDLVDYEEVLGKLGYYDAD	64
QY	94	GDGDFDVDDAKVLLGLKERSTSEPAVPBEAEPRTEPEQVPEABPQNIIDEAKEQIQS	153
Db	65	GDGDFDVDDAKVLLGLKERSPSERTFPF-EAETHAELEQAEQAGADIQNVEDVEVKEQIQS	123
QY	154	LHHEMWAHEHVEGEDLQOEDEGPTGEPOQODEFLMATVDVDRFETLEPEVSHETESHYH	213
Db	124	LLQESVHTDH---DL-EADGLAGEQPEVEDFLTVTDSDDRFEDLEPGTVEHEETEDTYH	178
QY	214	VEETVSQDCNQDMEEEMSEQENPDSSEHPVEDERLHHDITDVTYYQYEEQAVYEELENEG	273
Db	179	VEDTASQNHENDMEEMTNEQENSDBEAVTDAGVLLPHAEVYRHQDYDE-PVYEPSHEEG	237
QY	274	IEITEVTAPEDBNPVEDSQVIVEVSIFFPVEEQOEVPPEP	313
Db	238	VEIS-----DNTIDSSIISEEINVASVEEQQDTPPDT	270

RESULT 4

T18861
probable peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T18861, T23533
R/Swinburne, J.
submitted to the EMBL Data Library, August 1996
A/Reference number: Z19032
A/Accession: T18861
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-872 <WIL>
A/Cross-references: EMBL:Z79596; NID:el323798; PIDN:CAB01859.1; GSPDB:GN00028; CESP:K09A9
A/Experimental source: clone C02C6
R/Swinburne, J.
submitted to the EMBL Data Library, August 1996
A/Reference number: Z19753
A/Accession: T23533
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-872 <W12>
A/Cross-references: EMBL:Z79601; PIDN:CAB01887.1; GSPDB:GN00028; CESP:K09A9.6
A/Experimental source: clone K09A9
C/Genetics:
A/Gene: CESP:K09A9.6
A/Map position: X
A/Introns: 18/1; 40/3; 74/1; 104/2; 139/1; 170/3; 385/3; 432/2; 464/3; 530/2; 609/3; 776/3
C/Keywords: oxidoreductase

	Query Match	19.8%;	Score 798;	DB 2;	Length 872;	
	Best Local Similarity	26.8%;	Pred. No. 2e-33;			
	Matches 191;	Conservative 144;	Mismatches 278;	Indels 100;	Gaps 15	
QY	89 IYDADGDFDVDDAKVLTGLKERSTSEPAVPPEAEAPTEPEEQ-----VVEAEPQ	141				
Db	205 VEDDDDDDEDDDDVDVE-----APAQPESRKQKANHOREEKKDKNKKYQGVKEEPD	255				
QY	142 NIEDEAKEQIQSLHEMVHAHVEGEDLQOEDG-----PTGEPOQEDDEFMATDVDDR	195				
Db	256 LDDDDDDDEDDDDDDGGDDDDDEKDDGNDAEKDDGGDDDDDDDDGNDDEDEDGKNKKTSSVEAK	315				
QY	196 FETL-----EPEVSHEETE--HSYHVEET-----VSQDC-NODMEE-	228				
Db	316 SDKVVEEKNDGDDEPHVSHREARQLRQRHLRESRRDNRPQQGNRECIGHQCENKRESLKP	375				
QY	229 ---MMSEQENPDSSSEPVEDERLHHDITDVTYYOVEQAIFYPLENEGIEITEVTAPPE	284				
Db	376 RKSLLVTKKTKSMVERILDDPEFDDDDDD-----EDDSEYANK-----N	415				
QY	285 DNPVEDSQIVIVEVSIFFPVEEQGVPPETNRKTTDPQCAKVKKKKPKLLNFKDTIKAE	344				
Db	416 DREDDDDVDVERIS-----DDSSSSSYKRHAITTKKEIGFRDI	454				

Dz	345	LDAEKLRRKRGKIEEA VNA FKELV R KYPQSPRAR YGKA QCEDDLAEKRBSNEVLGAIET	404
Dz	445	LBRADNLVEKHQYEEAMELFDHVIAVYPA STRAYFGKARAYDIRGEI EADETRDKAIEI	514
Dz	405	YOEVASLPDVPADLLKLSLKRRSDROQLGHMRGSLTLQRLVQLFPNDTSLKNDDLGVGY	464
Dz	515	YEKILQNSGVDPDALFRQAORLI EKTRF RGQLAKHTLAHRYFIDRFPEELNLQTDFALISF	574
Dz	465	LLIGDNDAKKVYEEVLSTVPNDGPFAKVHYFILKA-QNKIAESIPYLKEGISGDEGTD	523
Dz	575	VMMKRYEDARTVLKNVLANDPNHVIALAYGYILKAHD DKEQVALMRKS LKNADNEIT	634
Dz	524	DGRFEYHLGDAMQRVG NK-EAYKMYELGHKRGFASFVWQRSLYNVNGLKAQPWWTPKETG	582
Dz	635	DKREYYQLGHGLTTLGKRS EADAVYQKAAQM GVEMTAOQRS LYNI EGLTGRAWAMDOTP	694
Dz	583	YTELVKSLERNWKLI RD EGI LAVMDKAGLLFPEDENLR EKGDW SQFTLMQQGRN ENACK	642
Dz	695	YSKFLKTV ERQWATIRQEGMEVLKD CSDCWL DHNQOLVIDGQWKFFPIMSEONFIKSSCE	754
Dz	643	GABKTCTLLEKP PETTGCRRGQIKYSIMHPGTHVMPHTGPTNCRLRMHLGLVIPKEGCKI	702
Dz	755	RMPOTCHILOEP AASSNA SKSMHL SVLSSGASILPHCGPTNYHLQAHGLIVSPSE-ARI	813
Dz	703	RCANETR TWBEGKVLIFDDSFHEHYWD--ASSFR LI FIVDVWHPELTPOOR	752
Dz	814	RVNGETKWRSGKF TI IYD DSFEHEHQFDGA SSSSFR LVLTIO LWHPEVQPHOR	866

RESULT 5

T47148
hypothetical protein DKFZp761P039.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47148
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24379
A;Accession: T47148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-186 <AAA>
A;Cross-references: EMBL:AL161993
A;Experimental source: adult amygdala; clone DKFZp761P039
C;Genetics:
A;Note: DKFZp761P039.1

```
Query Match      6.6%; Score 263.5; DB 2; Length 186;  
Best Local Similarity   35.4%; Pred. No. 6.5e-07;  
Matches    63; Conservative     28; Mismatches    80; Indels       7; Gaps        4;
```

DY VKSLERNMKLIRDEGLAVMDKAKGFLPE--DENTLRKGDSQFTLMQQGRNRNENACKGA 644
| : | | | : | : | : | : | : | : | :
Db VEVLERNFQTILCEPETLYKAFSNCSLPQGWMKNSTPSGEWFTFYLVNQGVCPRNCRKC 64

DY PKTCTLLLEKEFPETTGCR-RQQLKYSSIMHPGTHWPHPTGNRLRMHLGLVLPKEGGCKIR 703
| : | | | : | : | : | : | : | : | : | : | : | :
Db PRTYRLLGS LRTCIGNNVFGACISVLSPGVTITEHYGPINRI RCHLGKLTP-NGC ELV 123

DY CANETRTEEKKVLIFDSSFEEHVVQDASSF---RLFIIVDVWHPELLTPQQRSLPAI 758
| : | | | : | : | : | : | : | : | : | : | : | :
Db VEGEPQCWAEGRCCLFDSESELHAAPHEGSAEDGPRVENVDMLMHNPNVAALAEARQALDFI 181

RESULT 6

151116
NF-180 - sea lamprey
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: 151116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995


```
Qy      702 IRCANETRTWEEGKVLIFDDSFHEHVMQDASSFRLIFYDVWHP 745
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      185 IYVDGQPYAWRDGGEVAMEDEFVHWKNETEQTRVILFCDIERP 228
```

RESULT 9

zinc metalloproteinase ZmpB, probable [imported] - Streptococcus pneumoniae (strain TIGR
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: H95076
R/Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.U.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; EMID:11463916
A/Accession: H95076
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1881 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK74809.1; PID:gl4972138; GSPDB:GN00164; TIGR:SP4
A/Experimental source: strain TIGR4

Query Match	4.8%;	Score 191.5;	DB 2;	Length 1881;
Best Local Similarity	21.6%;	Pred. No. 0.06;		
Matches 138;	Conservative 82;	Mismatches 226;	Indels 193;	Gaps 30;

```

QY 111 ERSTSEPAVPEE--ABPHTPEBEOVPVEAEPOINIEDEAKEQOSSLHEMVAHAEHVGED 168
    : ::||| ||| : ::||| : : : : : | : |
Db 211 KEDSAEPA-PVEEVGGEVESKPEEKAVAPKESOPSCKPAEBSKVEQAAGEPV---APRED 265
    : : : : : : : : : : : : : : : : : :
QY 169 LOQEDGPTEGPQOEDEFIMATDVDRFETLEPEVSHEETSHSYAETVSODCNO-DME 227
    : : : : : : : : : : : : : : : : : :
Db 266 EKAPVEPEKQFEAPEER--KAHEETPKQEESTPDTKAEETVER--KEETVNQSIEQKVE 321
    : : : : : : : : : : : : : : : : : :
QY 228 EMMSSEQENPDSSSEPVEDE---RLHHDTDDVTQVVEEQAAVEPLENEGIEITEVTAPR 283
    : : : : : : : : : : : : : : : : : :
Db 322 TPBAVEKQTEPTTEEPKVEQAAGEPVAPAREDEQAPTAPVEPEKQPEVBEKAV---EEIPKP 378
    : : : : : : : : : : : : : : : : : :
QY 284 ED-----NPVEDSQV---IVEEVSIFVE-----EQQ 307
    || : : : : : : : : : : : : : : : :
Db 379 EEXIKIGITKEPVDKSELNNQIDKASSVSPIDYSTASYNALGPVLETAAGVYASEPYKOP 438
    : : : : : : : : : : : : : : : : : :
QY 308 EVPEPETNR-----KTQD-----DPEOKAKVKKKPKLKNFKDKTIKAEIDAEEK 350
    || : : : : : : : : : : : : : : : :
Db 439 EVNSETNKLTALDALNDVKTELNTLIAADAKTKVKEHYSDRSWONLOTEVTKAEKVYAANT 498
    : : : : : : : : : : : : : : : : : :
QY 351 LKRKGKIEEAVNAF-----KEIVRKYPQSPRARYGKAQ 383
    : : : : : : : : : : : : : : : : : :
Db 499 DAKOSEVNEAVEKLTATIEKLVELSEKPIILTLSTDKKILEREAVALYTLE--NQNKTK 555
    : : : : : : : : : : : : : : : : : :
QY 384 CEDDLAEKRRSNEVLRGAI-----ETYQ-----E 407
    : : : : : : : : : : : : : : : : : :
Db 556 IKSITAELKKGEEVINTVLLTDDKVTTETISAAFKNLEYKEYTLLSTMYYDRNGEETE 615
    : : : : : : : : : : : : : : : : : :
QY 408 VASLPDVPADLKLKLSLK--RRSDRQQFLGHMRGSLTLQLRVOLFPNDTSIKNLDLGVYL 465
    : : : : : : : : : : : : : : : : : :
Db 616 TLENQNIOQLDKKVELKNIKRTDLIKY----ENGKETNESLITITPPDDKSNY-----YL 665
    : : : : : : : : : : : : : : : : : :
QY 466 LIGDNNDN-----AKKYVEEVLSTPNDFGFAKHVGFIUK-AQNKIAES-IPYLKEGIES 517
    : : : : : : : : : : : : : : : : : :
Db 666 KITSNNOXTLLAVAKNIIEETTNGTPVYKYATAIADNLVSRTADNKFEBEYVHYIEK---- 721
    : : : : : : : : : : : : : : : : : :
QY 518 GDBGTDDGRFYFHLDGAMQRGVGNKEAYKYVELGHKRGHFASVWORSLYNV--NGLKAQPW 575
    : : : : : : : : : : : : : : : : : :
Db 722 --PKVHEDNVYNNFKEVLEAIQN-DPSKEYRLGQ-----SMSARNVWPNG---KSY 766
    : : : : : : : : : : : : : : : : : :

```

```

Db      | : ||      || : ||      | : |
767 ITKEFTG-----KLSSEGGKQFAITELEHPLF 793

```

RESULT 10

cytadherence-accessory protein (hmw1) homolog MG386 - Mycoplasma genitalium
C/Species: Mycoplasma genitalium
C/Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C/Accession: G64242
R/Praser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A/Title: The minimal gene complement of Mycoplasma genitalium.
A/Reference number: A64200; MUID:96026346; PMID:7569993
A/Accession: G64242
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1616 <TIGR>
A/Cross-references: GB:U39723; GB:L43967; NID:g1046092; PID:g1046097; TIGR:MG386
A/Experimental source: strain G-37
C/Genetics:
A/Genetic code: SGC3

Query Match	4.7%;	Score 189.5;	DB 2;	Length 1616;
Best Local Similarity	22.3%;	Pred. No. 0.063;		
Matches 142;	Conservative 78;	Mismatches 205;	Indels 213;	Gaps 35;

QY	75	FDLVDVYEVLGK-----LGIYD-----ADGDGFDDVDDAKVL-----	106
Dd	1091	FDTVKHAEAVFDKNQTQTGLEBPQVSSEAEVNDQTTTDTIVGEPEAVFVDVQPEKTTEVKED	1150
QY	107	-LGLKERSTSEPAVPPEEAPHTPEEQVPVEAEPONIDEAKEQIOSLHMEMVAHEVE	165
Dd	1151	DVENOQKVISEPQVEQQPGEAVFEPESAFAKFDSPVESVQDSQPPELVLEEVQTQPEIQPVE	1210
QY	166	GEDLQGEDPTGPGEPOQEDEFLMATDVDDRFEETLE-PEVSHETEHSYHVEETVSOD-CN	223
Dd	1211	SQPEATFD--TVQPEQTPOE---AKFDSPVETVEQPEFSSEPFTQQ--HVSEASFEDEPN	1262
QY	224	QDMEEWMSQENP--DS---SEPVEDERLHHDTDVTYQV-----YEEQAVERPL	269
Dd	1263	YDFDEPNYDFPDQPSYSDLQPSSEQYDVDEPNYDFDEPNYIEISKSEPFQFEPQVEQQP-	1321
QY	270	ENEGIEITEVTAPE-DNPVE---DSQ--VIVEEVSIFPEVEEQEVPPEPTRKRKTDDEQK	323
Dd	1322	--GEAVFEPESAFAKFDSPVESVQDSQPPELLBEVQTQPEIQPVESQPEATFDTVQPEQT	1378
QY	324	AKVKKKKPKLLNKFDKTIKAELDAEKLRKGKIEEAVNAFKELVRKYPPQSPRARYGAQ	383
Dd	1379	PQEA-----KFD-----SP-----	1387
QY	384	CEDDLAEKRRSNEVLRGAIETYQE--VASLPDVPADILLKLRSRDRQQLGHMRGSLL	441
Dd	1388	-----VETIQEPQVSSEPEV---VVQGENFEREKPE-----TVL	1417
QY	442	TLORLVOLEFP--NDTSLKNDLGVCYLILIGDNNDNAKYYEEVLSTPNDGFAKYHYGFI	498
Dd	1418	EEOADEIOPEAASEESLDWELLVG-----NNSYGHYEP-----DG-BMWAGFFG	1462
QY	499	KAQ-NKIAESIPLYKEGIESGDPGTDDGRFYFHL-GDAMO-RVGKAEAYKWYELGHKRG	554
Dd	1463	DDQKMNKDA-TVKWARE-----RDYLPFIGDEVYGRYNNKGEMITWYGFIYESG	1509
QY	555	HFASV---WORSLYNVNGLKAQPWTFP---KETGYTELKLSLERNMKLIRDEGLAVMDK	607
Dd	1510	DWVLVDEQWKNRQPRIN--EAPKFWEKLI GNBEYGYE---DNEMNW-----	1551
QY	608	AKGLFLPEDENIREKGDWSOFTLMQOGRRMENACKGAP	645
Dd	1552	-----YDEEFDSEGNNMLVFQSEETENLNEDITKDIP	1582

QY 161 AEHVEG--EDLQOEDGPTGPEQOEDD-----EFLMATDV-----DDEPFTLEPEVS 204
Db 92 VTKDEGQAEATNMDEDADGKKEQCTDDGVSVEDTVMKENVESKDNNYAKDEKETKETDIT 151
QY 205 HEETESHYHVEETVSQDCNQ-----DMEE---MMSQENPDSSPEVVEDERLHHD 251
Db 152 --EADHKKAGKEDIQHEADKANNGKDGNTGDIKEEGTLVDEDKGTDM-----DEKVENG 203
QY 252 TDDVTYQVYEEQAVYEPLNEGIEITTEVTAPPEDNPVEDSQVIVEVS----IFPVEEQ 307
Db 204 DENKQVENVEGKEKEDKEENKTEVEAKAEVDESKVEDEKEGSEDENNEKVESKDAKE 263
QY 308 EVPFETNRKTDDEPEQAKAVKKKKPKLNFDKTIKAELEDAEKLKRGKIEEAVNAFKEL 367
Db 264 DEKETNDOKDEDEKESKSGKRGK-----GTSSGKVREKNTTEV----- 305
QY 368 VRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAI--ETVQEV-----ASLPDVPADL 418
Db 306 --KKDABPRTPF-----SDRPVREKRSVERLVALIDKDSKEFRVEKRGAYLKIDIPNA 358
QY 419 LKLSIKRRSDRQOFL-----GHWRSGLTLQRLVOLFPNDTSLKNDLGV--GYLLIGDNDN 472
Db 359 NKVMRKRSDETLKLLHPLFGFRGKAQAQIK-----TNILGFSGFVWHGDEKK 406
QY 473 AKKYVEEVLVTPND---GFAKVHYGFILKAQNKLAESIPLYKEGIE-----SGDPGTDDG 525
Db 407 AKEKYEKLEKCTEKEKWEFCVDLDIHITKATTKKEDIITKLFEFLEKPHVTGDTVGD-- 464
QY 526 RFVFHGDAMQVRGNKEAKYMYELGHKRGHPASVWORSLYNVNGLKAQPMWTPKETGYTE 585
Db 465 -----TTVSEKEK---SSKGAKRRK-----TPKKTISPTA 490
QY 586 LVKSIERNMWKLIRDEGLAVMDKAKGLFLPEDENIREK 622
Db 491 GSSSKRSKSAKSQKKSEATKVVKSLASHDSESESEK 527

RESULT 14
A33430

h-caldesmon - chicken

N:Alternate names: caldesmon, smooth muscle; calmodulin- and actin-binding protein

C:Species: Gallus gallus (chicken)

C>Date: 27-Feb-1990 #sequence_revision 27-Jun-1994 #text_change 22-Jun-1999

C:Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022

R:Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.

Biochem. Biophys. Res. Commun. 164, 503-511, 1989

A>Title: Primary structure and functional expression of h-caldesmon complementary DNA.

A:Reference number: A33430; MUID:90026426; PMID:2803315

A:Accession: A33430

A:Molecule type: mRNA

A:Residues: 1-771 <HAY>

A:CROSS-references: GB:M28417; NID:g211895; PIDN:AAA48810.1; PID:g211896

A:Experimental source: gizzard

A>Note: part of this sequence was confirmed by protein sequencing

R:Bryan, J.; Imai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.

J. Biol. Chem. 264, 13873-13879, 1989

A>Title: Cloning and expression of a smooth muscle caldesmon.

A:Reference number: A32642; MUID:89340480; PMID:2760048

A:Accession: A32642

A:Molecule type: mRNA

A:Residues: 1-318,334-771 <BRY>

A:CROSS-references: GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:g212657

A>Note: the authors translated the codon GAA for residue 743 as Ieu

A>Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)

R:Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.

Biochem. Biophys. Res. Commun. 161, 38-45, 1989

A>Title: 35KDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo

A:Reference number: A32445; MUID:89273666; PMID:2730665

A:Accession: A32445

A:Molecule type: mRNA

A:Residues: 466-771 <HA2>

A:CROSS-references: GB:M26684; NID:g211897; PIDN:AAA48811.1; PID:g211898

R:Mac, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.

J. Biol. Chem. 266, 19971-19975, 1991

A>Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphoryl

A:Reference number: A41064; MUID:92041815; PMID:1939059

A:Accession: A41064

A:Molecule type: protein

A:Residues: 597-600;678-696;711-721 <MAX>

R:Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.

Biochem. Biophys. Res. Commun. 162, 746-752, 1989

A>Title: Caldesmon has two calmodulin-binding domains.

A:Reference number: A60461; MUID:89334885; PMID:2757638

A:Accession: A60461

A:Molecule type: protein

A:Residues: 2-17,'X',19-38;466-485 <WAN>

R:Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.

Biochem. Biophys. Res. Commun. 197, 145-153, 1993

A>Title: Common structural and expressional properties of vertebrate caldesmon genes.

A:Reference number: PC2003; MUID:94071934; PMID:8250919

A:Accession: PC2003

A:Molecule type: DNA

A:Residues: 74-419 <HAR>

R:Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.

J. Biochem. 106, 778-783, 1989

A>Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon

A:Reference number: PX0022; MUID:90130380; PMID:2613684

A:Accession: PX0022

A:Molecule type: protein

A:Residues: 462-477,'D',479-563;674-762,'A',763-771 <TAK>

C:Comment: This protein plays a vital role in the regulation of smooth muscle and nonmu

C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.

C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmo

C:Superfamily: caldesmon

C:Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; musc

F:1-771/Product: h-caldesmon #status predicted <HMAT>

F:1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>

F:266-390/Region: 13-residue repeats

F:511-582/Region: tropomyosin binding

F:622-636/Region: tropomyosin binding

F:597,682,717/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status experim

F:688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimenta

Query Match 4.5%; Score 181; DB 1; Length 771;
Best Local Similarity 22.8%; Pred. No. 0.066;
Matches 150; Conservative 87; Mismatches 248; Indels 174; Gaps 32;

QY 108 GLKERSTSE---PAVPPEE-----AEPHTPEEQV-----PVEAE-----PQNI 143
Db 178 GKKEEKDSEEEKPKVPTTEENQDVAVEKSTDKEEVETKTLAVNAENDTNAMLEGEQSI 237
QY 144 EDEAKEQIQSLHENVHAHVEGEDLQ-QEDGPTGEPQOEDDEFIMATVDVDRFETLPE 202
Db 238 TDAADKEKEAEKEREKLEAEKEKRLKAEEBKKAEEKQKAESEKKAABERERAKAEEK 297
QY 203 VSHEETESHYHVEETVSQDCNQDMEEMMSQENPDSSPEV-VEDERLHHDTDVTVQVYE 261
Db 298 RAAEERERAKAEE---RKAABERERAKAABERERAKAABEER-----KAAE 344
QY 262 EQAVYBPLENEGIEITTEVTAPPEDNPVEDSQVIVEVSIFPVEEQQEVPEETNR----- 315
Db 345 ERAKAE-----EERKAAEBERAKAEEKKAABERERAKAEESEKKAABEKLAEK 395
QY 316 --KTDDPEQAKAVKKKKPKLNFDKTIKAELEDA-----AEKL----- 351
Db 396 KEKKMEKKQAQEEKAQANLIRKQEBDEKAEVAKKESLPKLTQPTSKKDQVKNKDEK 455
QY 352 -----RKRGKIEE-AVNAFKELVKKYPQSPRARYGKAQCEDDLAEKRSNEVL 399
Db 456 APKEEMKSVWDRKRGVPEQKAQONGERELTTPKLTSTENAFGRSNLK-GAANAAGSEK 514
QY 400 GAETTYQVAVASLPDVPADLLKLSLKRSDRQOFLGHWRSGLTLQRLVOLFPNDTSLKND 459
Db 515 ---EKQOEAA---VELDELK--KRREERKIL-----EEBQKKK 546

QY 460 LGVGYLLIGDNDNAKKVVEV-----LSVTPNDFGA---KVHYGFIKKAQN-KIA 505
 Db 547 QEEAEKRIREEEKKRMKEIEIRRAEAEAERKQKVPEDGVSEKKPKCFSPKSSSLKIE 606
 QY 506 ESIPYLKESIESGDPGTDDGRFYFHLGDAMQVGNK-EAYKMYELGHRG----- 554
 Db 607 ERAEFLNKSAQK-----SGMKPAHTTAVVSKIDSRLEQYTSAVVGNKAAPAKPAASDL 660
 QY 555 -----HFASVQR-SLYNVNGLKAQPMWTPKET-GYTELVKSLERNWKLIRDEGLAV 604
 Db 661 FVPAEGVRNIKSMMEKGNVFSSPGGTGTP---NKETAGLKVGSSRINEMWLTKTPEG--- 714
 QY 605 MDKAKGLFLPEDENLREKGDWS-QFTLMQGRNRNENACKGAPKCTTLLEKFPETTGCR 662
 Db 715 -NKSPA---PKPSDLR-PGDVSGKRNLMWEKQVEKPAASSSKVTAT--GKKSETNGLRQ 766

RESULT 15

S47436
 flagellar antigen - Trypanosoma brucei (fragment)
 C/Species: Trypanosoma brucei
 C/Date: 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change 03-Nov-2000
 C/Accession: S47436
 R;Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T.
 submitted to the EMBL Data Library, August 1994
 A/Description: Repetitive proteins from the flagellar cytoskeleton of African Trypanosom
 A/Reference number: S47436
 A/Accession: S47436
 A/Molecule type: mRNA
 A/Residues: 1-411 <IMB>
 A/Cross-references: EMBL:Z36281; NID:G530358; PID:G530359
 A/Experimental source: strain stock TREU 1285
 C/Superfamily: cytoadherence-accessory protein hmw1

Query Match 4.5%; Score 179.5; DB 2; Length 411;
 Best Local Similarity 25.1%; Pred. No. 0.036;
 Matches 83; Conservative 53; Mismatches 128; Indels 67; Gaps 16;

QY 116 EPAVPEAEPEPTEPEEQVPEA-----EPQNIIDEAKEQIQSLHMHVHAB-HVEGEDL 169
 Db 11 EPQVPAAEQPEAQPEGDIAVEALEELEPEEQVPABAQP-----EAVAPEGDIAVEAL 63
 QY 170 QQEDGPTGEPQDEDEFLMATDVD-DRFETLE--PEVSHETEHSYHVEETVSQDCNQDM 226
 Db 64 EELPEPQAPAAEQPEALPEGDIAVEALEELEPEEQVPAAEQPEAVAPEGDIAVEALEEL 123
 QY 227 EEMMSEQENPDSSSP-VVEDERLHMDTDVTYQVVE-----QAVYEPLNEG--- 273
 Db 124 EE--PQVPAAEQPEAVAPE-----GDIAVEALEELEPEEQVPAAEQPEAVAPEGDIA 174
 QY 274 IEITEVTAPPEDNPE-----DSQVTVVEVSIFPVEEQEQVPEPTNRKTDDEQKAKV 326
 Db 175 VEALEELEPEEQVPAAEQPEAQPEGDIAVE--ALEELEPEEQVPAAEQPEAVAPEGDIAV 232
 QY 327 KKKPKLKNKFDKTIKAEIDAELKRGKIEEAVNAFKELVKKYPQSPRARYGKAQCED 386
 Db 233 -----EALAELEPEEQAPAAEQPEAQPEGDI--AVEALEEL-EPQQAPEAAEQPEAQPEG 284
 QY 387 DLAEKRRSNEVLGAIEITYQEVASLPDVPAD 417
 Db 285 DI-----AVEALEELEPEEQAPAE 303

Search completed: May 19, 2004, 16:00:36
 Job time : 34.7034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:51:04 ; Search time 19.2142 Seconds
(without alignments)
2054.168 Million cell updates/sec

Title: US-09-903-199-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSGSGS.....IVDVWHPBLTPQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3966.5	98.6	757	1 ASPH_HUMAN	Q12797 homo sapien
2	3172	78.9	754	1 ASPH_BOVIN	Q28056 bos taurus
3	203	5.0	700	1 TRDN_CANPA	P82179 canis fam1
4	194	4.8	4835	1 MDNI_GIALA	Q85t1 giardia lam
5	193	4.8	493	1 ECK1_METMA	Q8ptt8 methanosarc
6	189.5	4.7	1616	1 P200_MYCGE	Q49429 mycoplasma
7	187	4.6	705	1 TRDN_RABIT	Q28820 ocytolagus
8	181	4.5	771	1 CALD_CHICK	P12957 gallus gall
9	178.5	4.4	728	1 TRDN_HUMAN	Q13061 homo sapien
10	176.5	4.4	845	1 NFM_RAT	P12839 rattus norv
11	175	4.4	2004	1 MYG3_HUMAN	Q92794 homo sapien
12	172.5	4.3	1240	1 YNJI_YEAST	P53935 saccharomyc
13	172	4.3	465	1 YH06_YEAST	P38845 saccharomyc
14	172	4.3	630	1 YCF2_OENVI	P31569 cenothea v
15	169.5	4.2	592	1 LAM2_MOUSE	P21619 mus musculu
16	169.5	4.2	1233	1 YF16_YEAST	P43597 saccharomyc
17	165	4.1	650	1 NUCL_XENLA	P20397 xenopus lae
18	164	4.1	1957	1 SPOF_SCHPO	Q10411 schizosacch
19	163	4.1	411	1 FK83_YEAST	P38911 saccharomyc
20	162.5	4.0	793	1 CALD_HUMAN	Q05682 homo sapien
21	161.5	4.0	848	1 NFM_MOUSE	P08553 mus musculu
22	161.5	4.0	1220	1 IF2F_HUMAN	O60841 homo sapien
23	161.5	4.0	1395	1 SP41_YEAST	P38924 saccharomyc
24	161	4.0	2663	1 CENE_HUMAN	Q02224 homo sapien
25	161	4.0	4910	1 MDN1_YEAST	Q12019 saccharomyc
26	160.5	4.0	721	1 YCF2_OENPI	P31568 cenothea p
27	159	4.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
28	158.5	3.9	837	1 RA50_METTH	O26640 methanobact
29	158.5	3.9	1102	1 YG49_SCHPO	O60184 schizosacch
30	158	3.9	795	1 CDL1_HUMAN	P21127 homo sapien
31	157.5	3.9	780	1 CDL2_HUMAN	Q9ug88 homo sapien
32	157.5	3.9	1357	1 KINI_HUMAN	Q86up2 homo sapien
33	156.5	3.9	506	1 NPL3_HUMAN	Q99457 homo sapien

34	156.5	3.9	1658	1 YM67_YEAST	Q03661 saccharomyc
35	154.5	3.8	1828	1 MAP2_MOUSE	P20357 mus musculu
36	154	3.8	1549	1 TRHY_SHEEP	P22793 ovis aries
37	153.5	3.8	1898	1 TRHY_HUMAN	Q07283 homo sapien
38	152	3.8	434	1 YK12_YEAST	P36080 saccharomyc
39	152	3.8	671	1 CHS5_YEAST	Q12114 saccharomyc
40	152	3.8	725	1 HS9B_BRARE	O57521 brachydanio
41	152	3.8	728	1 HS9A_CHICK	P11501 gallus gall
42	152	3.8	1001	1 RPGR_MOUSE	P11501 mus musculu
43	152	3.8	1233	1 SM1A_BOVIN	O970x5 bos taurus
44	152	3.8	1233	1 SM1A_HUMAN	Q14683 homo sapien
45	152	3.8	1282	1 BMS1_HUMAN	Q14692 homo sapien

ALIGNMENTS

RESULT 1
ASPH_HUMAN
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-dioxygenase).
DE GN ASPH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95121937; PubMed=7821814;
RA Koriath F., Gieffers C., Frey J.;
RT "Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.";
RT Gene 150:395-399 (1994).
RL -1- FUNCTION: Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins.
CC -1- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC -1- COFACTOR: Iron.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
CC -1- TISSUE SPECIFICITY: Detected in all tissues tested.
CC -1- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 274-757) OR 52 kDa (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U03109; AAA82108.1; -.
CC PIR: I38423; I38423.
CC Genew; HGNC:757; ASPH.
CC MIM: 600582; -.
DR GO: GO:0005789; C:Endoplasmic reticulum membrane; TAS.
DR GO: GO:0005509; F:calcium ion binding; TAS.
DR GO: GO:0005489; F:electron transporter activity; TAS.
DR GO: GO:0004597; F:peptide-aspartate beta-dioxygenase activity; TAS.
DR GO: GO:0008307; F:structural constituent of muscle; TAS.
DR GO: GO:0006936; P:muscle contraction; TAS.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.

DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
KW Oxidoreductase; Dioxxygenase; Iron; Transmembrane; Signal-anchor;
KW Endoplasmic reticulum.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 1 54 LUMENAL (POTENTIAL).
FT DOMAIN 76 757 POLY-SER.
FT DOMAIN 13 20 POLY-LYS.
FT DOMAIN 323 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 757 AA; 85498 MW; 1A79313A4934C430 CRC64;

Query Match 98.6%; Score 3966.5; DB 1; Length 757;
Best Local Similarity 99.2%; Pred. No. 3.4e-193;
Matches 752; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 MAQRNNAKSSGSSSGSGSGSTAGSSPGARETKHGCHKGRKGLSGTSFTWEMY 60
Db 1 MAQRNNAKSSGSSSGSGSGSTAGSSPGARETKHGCHKGRKGLSGTSFTWEMY 60
QY 61 IALLGVMTSAVAVWFVFDLVYEBVLGKLTADAGDGFVDVDAKYLLGLKERSTSEAVP 120
Db 61 IALLGVMTSAVAVWFVFDLVYEBVLGKLTADAGDGFVDVDAKYLLGLKERSTSEAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEPONIDEAKEIQSLHEMVAHEVGEEDLQOEDGPTGEPC 180
Db 121 PEEAEPHTEPEEQVPVEAEPONIDEAKEIQSLHEMVAHEVGEEDLQOEDGPTGEPC 180
QY 181 QEDDEFMATVDVDRFETLREPEVSHETEHSHVEETVSQDCNQDMEMSEQENDSSSE 240
Db 181 QEDDEFMATVDVDRFETLREPEVSHETEHSHVEETVSQDCNQDMEMSEQENDSSSE 240
QY 241 PVEDERLHHDVDTVYQVYEEQAVYEPIENEGIEITEVTAPEDNPVEDSQVIVEEVS 300
Db 241 PVEDERLHHDVDTVYQVYEEQAVYEPIENEGIEITEVTAPEDNPVEDSQVIVEEVS 300
QY 301 FPVEEQEQEVPPEPTNRKTDDEQAKAVKKPKLNFDTIKAEILDAAEKLRRKGTIEA 360
Db 301 FPVEEQEQEVPPEPTNRKTDDEQAKAVKKPKLNFDTIKAEILDAAEKLRRKGTIEA 360
QY 361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRSNEVLRGATEYQEVASLPDVPADLLK 420
Db 361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRSNEVLRGATEYQEVASLPDVPADLLK 420
QY 421 LSLKRRSDRQQLGHRGSLTLQRLVQLFENDTSLKNDLGVGILLIGDNDNAKYYEEV 480
Db 421 LSLKRRSDRQQLGHRGSLTLQRLVQLFENDTSLKNDLGVGILLIGDNDNAKYYEEV 480
QY 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFHLGDAMQRYGN 540
Db 481 LSVTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFHLGDAMQRYGN 540
QY 541 KEAYKMYELGHRGHPASVWQSLYVNGLKAQPMWTPKETGYTELKSLERNMKLIRDE 600
Db 541 KEAYKMYELGHRGHPASVWQSLYVNGLKAQPMWTPKETGYTELKSLERNMKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRRENENACKAPKCTLLKEKPEPTGC 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRRENENACKAPKCTLLKEKPEPTGC 660
QY 661 RRGQIKYSIMHPTGTHWPHGTPTNCRLEMLGLVTPKEGCKIRCANETRTWEEGKYLIFD 720
Db 661 RRGQIKYSIMHPTGTHWPHGTPTNCRLEMLGLVTPKEGCKIRCANETRTWEEGKYLIFD 720
QY 721 DSFEHEVWQDAGSFRLLIFIVDVHPELTPQQRSLPAI 758
Db 721 DSFEHEVWQDAGSFRLLIFIVDVHPELTPQQRSLPAI 758

RESULT 2
ASPH BOVIN
ID ASPH BOVIN STANDARD; PRT; 754 AA.
AC Q28056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (Asp beta-hydroxylase) (Peptide-aspartate beta-dioxygenase).
GN ASPH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Liver;
RX MEDLINE=92332546; PubMed=1378441;
RA Jia S., Vandusen W.J., Diehl R.E., Kohl N.E., Dixon R.A.F., Elliston K.O., Stern A.M., Friedman P.A.;
RT "cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.";
RL J. Biol. Chem. 267:14322-14327(1992).
RN [2]
RP SEQUENCE OF 289-385 AND 615-641.
RC TISSUE=Liver;
RX MEDLINE=91310689; PubMed=1856229;
RA Wang Q., Vandusen W.J., Petrocki C.J., Garsky V.M., Stern A.M., Friedman P.A.;
RT "Bovine liver aspartyl beta-hydroxylase. Purification and characterization.";
RL J. Biol. Chem. 266:14004-14010(1991).
CC -1- FUNCTION: Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins.
CC -1- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC -1- COFACTOR: Iron.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
CC -1- PTM: Might be processed to the 56 kDa (AA 289-754) or 52 kDa (AA 311-754) forms in the lumen of the endoplasmic reticulum.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91213; AAA03563.1; -.
DR PIR; A42969; BABOH.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
KW Oxidoreductase; Dioxxygenase; Iron; Transmembrane; Signal-anchor;
KW Endoplasmic reticulum.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 58 78 LUMENAL (POTENTIAL).
FT DOMAIN 79 754 POLY-GLY.
FT DOMAIN 9 12 POLY-SER.
FT DOMAIN 14 21 POLY-LYS.
FT DOMAIN 318 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 754 AA; 84998 MW; 369593A1F0B558C8 CRC64;

Query Match
Best Local Similarity 78.9%; Score 3172; DB 1; Length 754;
Matches 612; Conservative 45; Mismatches 77; Indels 44; Gaps 6;

QY 1 MAQRKNKAKSGG---NSSSSSGSGSGSTASAGSSSPGARRETHGKHNGRKGGLSGTSFETW 57
Db 1 MAPRKNKAKGGGSSSSSSSSSGSPGTGCTSGSSSPGARRETHGKHNGRKGGLSGTSFETW 60
QY 58 FMVIALLGWTSVAVWFDLVYEEV-----GKLGIYDADGDGDFDVID 102
Db 61 FMVIALLGWTSVAVWFDLVYEEVLAKAADFVNLSEVLQGLGIYDADGDGDFDVID 120
QY 103 AKVLGLKERSTSEPAVPPPEAEPEHTEPEEQVPEAEAPONIEDEAKEQIQSLHEMVHAE 162
Db 121 AKVLGLKERKPAKPTVPPEEADMYFWLEDQVLESPPRONIEDEVYEQVQS-LDETYSSE 179
QY 163 HVEGEDIQEDGPTGEPQOEDEFLMATDVDDRFETLEPEVSHETESHVHEETVSQDC 222
Db 180 --PGENLPQEPPEGPABELQPDHVFVGSADADDRYEPMGTCVAHSETEDSYHIEETASPAY 237
QY 223 NQDMEEMSGENPDSSSEPVVED--ERLHHDTDVYQVYEQAVYEPLNEGIEITEVT 280
Db 238 SQDMEEMWYEQENPDSSSEPVVDAERTYQETDDVYRDYDEQ----- 280
QY 281 APPEDNPVEDSQVIVEEVSIFVEHQEQVPPETNRKTDDEQAKAVKKKKPKLNFKDKT 340
Db 281 ----DHAVDNSNTLIEEPHMPRAEQEQVPPETNRKADDEPGKKKKKKPKLNFKDKT 336
QY 341 IKAELDAAEKLRRKGIIEAANAFAKELVRKYPOSPPARYGAQCEDDLAEKRSSNEVLRG 400
Db 337 IKAELDAAEKLRRKGIIEAANAFAELVRKYPOSPPARYGAQCEDDLAEKRSSNEILRR 396
QY 401 AIETTYQEVASLPDVPADLLKLSIKRRSDRQOFLGHMRSLLTLQRLVOLFPNDTSLKNDL 460
Db 397 AIETTYQEAASLPDAPTDLVKLSIKRRSDRQOFLGHMRSLLTLQRLVOLFPNDTALKNDL 456
QY 461 GVGYLLIGDNDNAKVVYEEVLSYTPNDGFAKVHYGFIKAQNKAIESIPYLKEGIESGDP 520
Db 457 GVGYLLIGDNDNAKVVYEEVLSYTPNDGFAKVHYGFIKAQNKAIESIPYLKEGIESGDP 516
QY 521 GTDDGRFVYHLDGAMQVRVGNKEAYKMYELGHKRGHFAVWQSRSLYNVNGLKAQPMWTPKE 580
Db 517 GTDDGRFVYHLDGAMQVRVGNKEAYKMYELGHKRGHFAVWQSRSLYNVNGLKAQPMWTPKE 576
QY 581 TGYTELVSLSERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRRNENA 640
Db 577 TGYTELVSLSERNWKLIRDEGLAAMDRTHTGLFLPEDENLREKGDWSQFTLMQGRKNENA 636
QY 641 CKGAPKTCTLLEKPEPTTGCRGOIKYSIMHPTGTHVWPHGTPTNCRRLRMHLGIVIPKEGC 700
Db 637 CKGAPKTCSLLDKPETTGCRGOIKYSIMHPTGTHVWPHGTPTNCRRLRMHLGIVIPKEGC 696
QY 701 KIRCANEETRTWEEGKVLIFDSEFEHEVWQDASSFRLIFVDVWHPHPELTPOQRSLPAI 758
Db 697 KIRCANEETRTWEEGKVLIFDSEFEHEVWQDASSFRLIFVDVWHPHPELTPOQRSLPAI 754

RESULT 3
TRDN CANFA STANDARD; PRT; 700 AA.
AC P82179;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triadin.
GN TRDN.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;

RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Heart, and Skeletal muscle;
RX MEDLINE=99428545; PubMed=10497235;
RA Kobayashi Y.M., Jones L.R.;
RT Identification of triadin 1 as the predominant triadin isoform
RT expressed in mammalian myocardium.";
RL J. Biol. Chem. 274:28660-28668(1999).
CC -1- FUNCTION: May be involved in anchoring calsequestrin to the
CC junctional sarcoplasmic reticulum and allowing its functional
CC coupling with the ryanodine receptor (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=Skeletal;
CC IsoId=P82179-1; Sequence=Displayed;
CC Name=Cardiac 1;
CC IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
CC Name=Cardiac 3;
CC IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC -1- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF165916; AAF00222.1; -.
DR EMBL; AF165915; AAF00221.1; -.
DR EMBL; AF165917; AAF00223.1; -.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 47 67 POTENTIAL.
FT DOMAIN 68 700 LUMENAL (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 257 277 DOYAFCRYMIDMFVHGDLPFG -> GKHSSEVAGSKRTLG
FT KKQIQ (in isoform Cardiac 1).
FT /FTId=VSP_004001.
FT /FTId=VSP 004002.
FT VARSPLIC 278 700 Missing (in isoform Cardiac 1).
FT /FTId=VSP 004002.
FT VARSPLIC 466 466 E -> EPIKGEVAVPESLKEKE (in isoform
FT Cardiac 3).
FT /FTId=VSP_004003.
FT VARSPLIC 524 579 EEKVVQVKAETKAIEKTVKPKPAKKAHEOKESPITKD
FT KPKPTSKETPEVETES -> GILQVVPVNLCLFLVQFOQDE
FT ELNVEKVFVMIHVLHSHPTSRTSPLVISTGCT (in
FT isoform Cardiac 3).
FT /FTId=VSP_004004.
SQ SEQUENCE 700 AA; 78152 MW; F033E3AA1BEE0C56 CRC64;

Query Match
Best Local Similarity 5.0%; Score 203; DB 1; Length 700;
Matches 141; Conservative 95; Mismatches 236; Indels 258; Gaps 28;

QY 6 NAKSSGNSSSGSGSGSTAGS-----SSPGARETHGKHNGRKGGLSG 51
Db 8 NASTTTVIDSKNGSVKPGKVLKRTVTEDIVTTFSSPA----- 48
QY 52 TSFTFMVIALLGWTSVAVWFDLVYEEV-----LGKLG-----IYDA----- 92
Db 49 -----WLVIALIITWSAVAVWFDLVYKNFSSASSLSKIGSDPLKLVHDAVEETDWWY 103
QY 93 -----DGGDPDFVDVDAKVLGLKERSTSEPAVPPPEAEPEHTEPEEQV 135
Db 104 GPFSLSDIISDGEDDDDDDDTDKGEI-----EPPPLKQKEIHKEKAEKEKPERKITL 159

QY 136 VEA---EPONIEDAKEQIQLHMEVHAHEVEGEDLQOEDPTGEPQOEDDEFIMATDV 192
Db 160 AKVAHREKEKVEKEKSE-----KKATHKEKIEKKEKPEKTKMAKEERKAKTEEKIKKEV 214
QY 193 -DDRFETLEPEVS-----HEETEH-----SYHVEETVSQDC 222
Db 215 KGGQKEKVKPTAAKVEVQKTPPKAKEKEGKETA AVAKHEQKQDYAFCRYMIDMFVHGDL 274
QY 223 NQDMEEM-----SEQENPDSSPEVVEDERLHHTDDVTYQYVE 261
Db 275 RFGQSPALPPLPTVQASRPTPASPTLEGEKEEKKAEKKTSETKKKEKEDVKKKSDK 334
QY 262 BOAV-----YEPLENEG-IEITEVTAPEDNPVEDSQVIYEVSIFPVEEQOE----- 308
Db 335 DTAIDVEKKEPKAPETKQGTIKVAQAAKKDEKEDSKTKTPVEEHPKGGKQEKKEK 394
QY 309 -VPEETNRKTTD--PEQK-----AKVKKKKP-KLINKDKTIKAEILDAAE 349
Db 395 YVEPAKSSKKEHSAPSEKQVKAKTERAKETSAASTKKA VPGKKEKTKTVEQEI----- 450
QY 350 KLRKRGKI-----EAVNAFKELVRKYPOSPPARYGK-----AQCEDD 387
Db 451 RKEKSGKTSTASKDKEPEIKKDEKMPKADKVKPKPPQSQVKKKESESQVKKAEKPEQD 510
QY 388 LAEKRS-----NEVLGATF-----TYQVYASLPDVPADLKLKLS 422
Db 511 IAKPEKTVSHGKPEEKVVKQVKA TEKAIEKTVKPKPAKKAHEQEKES-PTIKTDKPKET 569
QY 423 LK-----RRSDRQOFLGHNGSSLTLQRLVQLFPNDTSLKNDLGVGYL 465
Db 570 SKETPEVTESGKKKIEKSEKSEKAEKMKLKEBKVSTRKESTLQSHNVTKAEKPARVSR 629
QY 466 LIGDNDNAKKVEEVSIV-----TPNDGFAKVH-----YGFILKAQNKIAESIPYIK 512
Db 630 DLEDVSASKKAKEAEADVSTTKRQKSPISFPQCVYLLDGYNGYGFQFPV-----TPAYR 682
QY 513 EGIESGDPGT 522
Db 683 PGESSGQPPS 692

RESULT 4
MDN1_GIALA STANDARD; PRT; 4835 AA.
ID MDN1_GIALA
AC Q8T5T1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Midasin (MIDAS-containing protein).
GN MDN1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20389616; PubMed=10930750;
RA McArthur A.G., Morrison H.G., Nixon J.E., Passamaneck N.Q., Kim U.,
Hinkle G., Crocker M.K., Holder M.E., Farr R., Reich C.I., Olsen G.E.,
RA Aley S.B., Adam R.D., Gillin F.D., Sogin M.L.;
RT "The Giardia genome project database."
RL FEMS Microbiol. Lett. 189:271-273 (2000).
RN [2]
RP IDENTIFICATION, GENE NAME, AND SIMILARITY WITH OTHER FAMILY MEMBERS.
RX PubMed=12102729;
RA Garbarino J.E., Gibbons I.R.;
RT "Expression and genomic analysis of midasin, a novel and highly
RT conserved AAA protein distantly related to dynein."
RL BMC Genomics 3:18-18 (2002).
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC the assembly/disassembly of macromolecular complexes in the
CC nucleus.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 VFMA domain.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF94287; AAM12656.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003599; AAA_ATPase_centre.
DR InterPro; IPR002035; VFMA_A.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 5.
DR SMART; SM00327; VFMA; 1.
DR PROSITE; PS50234; VFMA; 1.
KW Chaperone; ATP-binding; Repeat; Nuclear protein.
FT NP_BIND 356 363 ATP (POTENTIAL).
FT NP_BIND 814 821 ATP (POTENTIAL).
FT NP_BIND 1127 1134 ATP (POTENTIAL).
FT NP_BIND 1513 1520 ATP (POTENTIAL).
FT NP_BIND 1839 1846 ATP (POTENTIAL).
FT NP_BIND 3277 3284 ATP (POTENTIAL).
FT DOMAIN 110 113 POLY-LEU.
FT DOMAIN 3004 3007 POLY-LEU.
FT DOMAIN 4156 4161 POLY-ASN.
FT DOMAIN 4629 4818 VFMA.
SQ SEQUENCE 4835 AA; 539726 MW; 3A9E12417DB04A50 CRC64;

Query Match 4.8%; Score 194; DB 1; Length 4835;
Best Local Similarity 23.1%; Pred. No. 0.084;
Matches 119; Conservative 74; Mismatches 203; Indels 120; Gaps 25;

QY 80 YEEVLGKLGIVD--ADGDGDFVDVDAKVLGLKERSTSEPAVPPEAEHPTEPEEQVYE 137
Db 4188 HEEQADATGSTEQAQEDDYNDLDD-KNLISG-----QSDLSVPEADGEDEVNEE--LE 4238
QY 138 AEPONIEDAKEQIQLHMEVHAHEVEGEDLQOEDPTGEPQOEDDEFIMATDVDDRE 197
Db 4239 EEOQOMSDLSNPD-----QDACAIEEDDDRDRLPSSD-----ENAEHDEHEAPVDIDDN-E 4288
QY 198 TLPEVSHETESHYVEETVSQDCNODMEEMSEOE-----NPDSEPVVEDER- 247
Db 4289 ASDEQSTYNDNRDDAINISAQQQATNDEEEMQKTEYDQENITDSNPDANEVGTNDQKQ 4348
QY 248 LHHDTDDVTYQYVEEQAVYEPLEN--EGITEVTAPEDNPVEDSQVI-VEEVSIFPV 303
Db 4349 THEDNDQFRQENIEDQWEAESESTENSGEAGSADLKEGNPMSLEEFQRIWKERLNIHDR 4408
QY 304 E-----EQQEVPPETNRKTTDDEPKA-----KVKKKPKLNLKFDKTIKAEILDAAE 349
Db 4409 ESEKDEAAEPQDMPLOSNKTVPEFDDSKSGRDGALGLTESKHNLTNQ-----EFDNPN 4461
QY 350 KLRKRGKIEAVNAFKELVRKYPOSPPARYGKAQCEDDLAE-----KRR 393
Db 4462 EER--NVEH--NSSCETSQSHDRPPAEHLNPEISDEGESSTASDKQEQAVLSHMEBS 4516
QY 394 SNEVLGATETQOEVA-SLPD-----VPADL-----IKLSLKRSDRQOFLGHMRG 438
Db 4517 SKDLNPEGEVYQELAVSLASEETKRAPEDVAASARGNHLDDLKQTSAAAF----- 4570
QY 439 SLTLQRLVQLFPNDTS-LKNDLGVGYL-----IGDNDNAKKVYE 478
Db 4571 SLAERLRIT-LEPTVTSDLKGFRTGKTLNLRRIIPFIASEPQDKIWLRTKPSKRYVQ 4629
QY 479 EVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEG 514
Db 4630 VLLAVDDSSSMAPV-AKYALQATLLEFNACKFLBVG 4664

RESULT 5
ECX1_METMA


```

ID  ECX1 METWA      STANDARD;      PRT;      493 AA.
AC  Q8PT8;
DT  10-OCT-2003 (Rel. 42, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN  MM2623.
OS  Methanosarcina mazel (Methanosarcina frisia).
OC  Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC  Methanosarcinaceae; Methanosarcina.
OX  NCBI_TaxID=2209;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX  MEDLINE=22120827; PubMed=12125824;
RA  Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA  Martinez-Arias R., Henne A., Wierzer A., Baumer S., Jacobi C.,
RA  Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA  Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA  Fritz H.-J., Gottschalk G.;
RT  "The genome of Methanosarcina mazel: evidence for lateral gene
RT  transfer between Bacteria and Archaea.";
RL  J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
CC  -1- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC  of RNA species (Potential).
CC  -1- SUBUNIT: Component of the archaeal exosome multienzyme
CC  ribonuclease complex (Potential).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC  -1- SIMILARITY: Belongs to the RNase PH family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE013507; AAM32319.1; -.
DR  HAMAP; MF_00591; -; 1.
DR  InterPro; IPR001247; 3_ExoRNase.
DR  Pfam; PF01138; RNase_PH; 1.
DR  Pfam; PF03725; RNase_PH_C; 1.
KM  Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
FT  DOMAIN 1 254 PROBABLE EXOSOME COMPLEX EXONUCLEASE 1.
FT  DOMAIN 255 493 UNKNOWN.
SQ  SEQUENCE 493 AA; 55248 MW; 979A757BEF8DC090 CRC64;

Query Match          4.8%; Score 193; DB 1; Length 493;
Best Local Similarity 26.3%; Pred. No. 0.0056;
Matches 97; Conservative 45; Mismatches 129; Indels 98; Gaps 17;

QY  61 IALLGWTSAVAVFPLVDYEEVLKGLIYDADGDGDFV---DDAKVLL----- 107
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  153 IPMKGLITSCA---FGKVDGKIVLDLNKEDNYGEADFPVAMTQDGEITLIQMDGNLTPD 209
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  108 -----GLKERSTSEPAV-----PPEEAEPHTPE-----EQVPEAEPQ 141
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  210 EIKQGLELVKKGCKEILLEIQQAVLRKKFETPVEBVSSEETAPDKAEKVLPSVAIVE 269
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  142 NIDEAKEQIQSLIHEMVHAEHVEGDLQEDGPTGEPQED--DEFLMAT--DYDDRFE 197
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  270 ETPEEAEP-----EVEISEEVEAETILASEVTPDFEDLPEEIEELSESEEDLSETEE 323
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  198 TLEPEVSHETESHVETVSQDCNQDMEEMSEQENPDSSEPVDE----- 246
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  324 EFEEFALPEEAPEEDLEEDLEEDLGELEEEELPEEPEEALEETELASDECAP 383
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  247 -----RLHHTDDVTYQVVE-----EQAVVEPLENEGIEITEVTAPPEDNVED 290
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  384 ELKEFDEIEARL--EKEDASIEAEETITPEAEAEATEEGLEEEA-EIEETAASEEN-IEA 439
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  291 SQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQAKAKVKKKKKRLINKPKTIKAILDAAEK 350
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
  
```

```

Db  440 EABAEEAE--PEVEAEETSTEAEEAEPEEE-----KSEGPW-----KVVDPSAG-- 486
QY  351 LRRKGKIEE 359
    | : | |
Db  487 --TRGEKDE 493

RESULT 6
P200 MYCGE      STANDARD;      PRT;      1616 AA.
AC  Q49429; Q49259; Q49298; Q49352; Q49353;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Protein P200.
GN  MG386.
OS  Mycoplasma genitalium.
OC  Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX  NCBI_TaxID=2097;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 33530 / G-37;
RX  MEDLINE=96026346; PubMed=7569993;
RA  Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA  Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA  Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA  Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA  Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA  Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT  "The minimal gene complement of Mycoplasma genitalium.";
RL  Science 270:397-403 (1995).
RN  [2]
RP  SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC  STRAIN=ATCC 33530 / G-37;
RX  MEDLINE=94075230; PubMed=8253680;
RA  Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III,
RT  "A survey of the Mycoplasma genitalium genome by using random
RT  sequencing.";
RL  J. Bacteriol. 175:7918-7930 (1993).
CC  -1- FUNCTION: COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN
CC  CYTADHERENCE (BY SIMILARITY).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U39720; AAC71613.1; -.
DR  EMBL; U02245; AAA03400.1; -.
DR  EMBL; U02245; -; NOT_ANNOTATED_CDS.
DR  EMBL; U02175; AAD12458.1; -.
DR  EMBL; U02126; AAD12402.1; -.
DR  PIR; G64242; G64242.
DR  TIGR; MG386; -.
KM  Cytadherence; Structural protein; Repeat; Complete proteome.
FT  DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT  REPEAT 1205 1236 1-1.
FT  REPEAT 1358 1389 1-2.
FT  DOMAIN 891 1389 2 X 26 AA REPEAT.
FT  REPEAT 1161 1186 2-1.
FT  REPEAT 1310 1339 2-2.
FT  REPEAT 256 256 P -> S (IN REF. 2).
FT  CONFLICT 304 304 S -> F (IN REF. 2).
SQ  SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;

Query Match          4.7%; Score 189.5; DB 1; Length 1616;
Best Local Similarity 22.3%; Pred. No. 0.037;
Matches 142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;
  
```

```

QY 75 FDLVDYEVLGK-----LGIVD-----ADGDGDFDVAKVL----- 106
Db 1091 FDTYKHEAVFDKNQQTGLEBPQVSSEAEVVDQTTTDTVGEPFAVFDVQPEKTEVKFD 1150
QY 107 -LGKERSTSEPAVPEPAEPHTEPEEQVPAEAPQNIIDEAKQIQSLHMHVAEHYE 165
Db 1151 DVENQOKVISEPQVEQGEAVFEPSSAEAKFDSVSVQDSQPEVLEEVQTOPEIQPVE 1210
QY 166 GEDLQOEDEPTGEPEQDEDFLMTADVDDREPTLE-PEVSHETEHSYHVEVTSQD-CN 223
Db 1211 SOPEATPD-TVQPEQTPQE---AKFDSPVETVEQPEFSSEPTQO--HVESEASFDEPN 1262
QY 224 QDMEEMMSEQENP--DS---SEPVEDERLHHTDDVTYQV-----YEEQAVYEPL 269
Db 1263 YDFDEPNYDFDQPSYSDLQPSSEPQYDVDEPNYDFDEPNYEIESKPEPQPEPQVEQGP- 1321
QY 270 ENEGIEITEVTAPPE-DNPVE---DSQ--VIVEEVSIFPVEEQGEVPEPNKTDDEQK 323
Db 1322 ---GEAVFEPSSAEAKFDSVSVQDSQPEPLLEEVQTOPEIQPVESQPEATFDTVQPEQT 1378
QY 324 AKVKKKKPKLNLNKEDTKIKALDAEKLRRKGKIEBAVNAFKELVRYKQSPRARVYKQ 383
Db 1379 PQEA-----KFD-----SP----- 1387
QY 384 CEDDLAEKRRSNEVLRGAIETYQE--VASLDPVPADLLKLSTKRSDRQOFLGHMRGSL 441
Db 1388 -----VETIQEPQVSSPEBV---VQPNFERKPE-----TVL 1417
QY 442 TLQRLVOLFP--NDTSLKNDLGVGYLLIGDNDNAKKVYEVLSTPNDFGFAKVHYGFL 498
Db 1418 EEPQADEIQPEASEESLDWELLVG-----NNSYGHYEP-----DG-EWVWAGPFG 1462
QY 499 KAQ--NKIAESIPYLKEGIESGDPGTDDGRFFHL-GDAMQ-RVGNKEAYKYELGHRKG 554
Db 1463 DDQKMNKA-TVKWARE-----RDYLLPLIGDEVYGRYNNKGEMIWYGFYDESG 1509
QY 555 HFAVSV--WQSLVNVNGLKAQFPWTP-----KETGYTELKSLERNWKLIRDEGLAVMDK 607
Db 1510 DMVLVDEQWKNRQPRIN--EAPKFWKELIGNEEYGYE---DNEMNW----- 1551
QY 608 AKGLFLPEDENLREKGDWSQFTLMQOGRRENACKGAP 645
Db 1552 -----YDGEFDSGNWLVFQSEETENLNEDITKDIP 1582

RESULT 7
TRDN RABIT
ID TRDN RABIT STANDARD; PRT; 705 AA.
AC Q28820; Q28636; Q28637; Q28643;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
RX MEDLINE=93286104; PubMed=7685347;
RA Knudson C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
RT "Primary structure and topological analysis of a skeletal muscle-
specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL J. Biol. Chem. 268:12646-12654 (1993).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC TISSUE=Skeletal muscle;
RX MEDLINE=94298946; PubMed=8026576;
RA Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;
RT "Structural diversity of triadin in skeletal muscle and evidence of
RT its existence in heart.";
```

```

RL FEBS Lett. 348:17-20 (1994).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RC TISSUE=Heart muscle;
RX MEDLINE=96132942; PubMed=8550602;
RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT "Biochemical characterization and molecular cloning of cardiac
RT triadin.";
RL J. Biol. Chem. 271:458-465 (1996).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX MEDLINE=96066664; PubMed=7578102;
RA Fan H., Brandt N.R., Caswell A.H.;
RT "Disulfide bonds, N-glycosylation and transmembrane topology of
RT skeletal muscle triadin.";
RL Biochemistry 34:14902-14908 (1995).
CC -1- FUNCTION: May be involved in anchoring calsequestrin to the
CC junctional sarcoplasmic reticulum and allowing its functional
CC coupling with the ryanodine receptor.
CC -1- SUBUNIT: Homooligomer of variable subunit number; disulfide-
CC linked.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Comment=Additional isoforms seem to exist;
CC Name=Skeletal 1; Synonyms=ST1;
CC IsoId=Q28820-1; Sequence=Displayed;
CC Name=Cardiac 1; Synonyms=CT1;
CC IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
CC Name=Cardiac 2; Synonyms=CT2;
CC IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
CC Name=Cardiac 3; Synonyms=CT3;
CC IsoId=Q28820-4; Sequence=VSP_004466;
CC Name=Skeletal 2; Synonyms=ST2;
CC IsoId=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
CC VSP_004465;
CC Name=Skeletal 3; Synonyms=ST3;
CC IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC -1- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31540; AAC48496.1; -
DR EMBL; U31540; AAC48496.1; -
DR EMBL; U31555; AAC48497.1; -
DR EMBL; U34201; AAC48498.1; -
DR PIR; A45990; A45990.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET 0
FT DOMAIN 1 46
FT TRANSMEM 47 67
FT DOMAIN 68 705
FT CARBOHYD 74 74
FT CARBOHYD 624 624
FT VARSPLIC 264 285
FT VARSPLIC 264 307
FT VARSPLIC 286 705
FT VARSPLIC 308 705
CYTOPLASMIC.
POTENTIAL.
LUMENAL.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .)
DOYAFRCYMDIDIFVHGDLKPGSPAIIPSPTEQASRPTRA
LPT -> ECIFLSAATPQGITNRQQLNDIHHCFLKTKKGGN
GOHAFCLKGC (in isoform Cardiac 2).
/FTId=VSP_004459.
Missing (in isoform Cardiac 1).
/FTId=VSP_004460.
Missing (in isoform Cardiac 2).
/FTId=VSP_004461.
```

FT VARSPPLIC 415 415 A -> E (in isoform Skeletal 2).
 FT VARSPPLIC 416 424 /FTId=VSP_004462.
 FT VARSPPLIC 584 584 Missing (in isoform Skeletal 2).
 FT VARSPPLIC 584 584 /FTId=VSP_004463.
 FT VARSPPLIC 584 584 D -> E (in isoform Skeletal 2 and isoform Skeletal 3).
 FT VARSPPLIC 585 591 /FTId=VSP_004464.
 FT VARSPPLIC 585 591 Missing (in isoform Skeletal 2 and isoform Skeletal 3).
 FT VARSPPLIC 645 705 /FTId=VSP_004465.
 FT VARSPPLIC 645 705 SKRAKEAEBSSTKXKSPISFQCVYLDGYNGYGFQFPV
 FT VARSPPLIC 645 705 TPAQYPGESSGKPNPSPKQ -> LATVGWGMNQMEDL
 FT VARSPPLIC 645 705 SVTLPSK (in isoform Cardiac 3).
 FT VARSPPLIC 645 705 /FTId=VSP_004466.
 SQ SEQUENCE 705 AA; 79003 MW; 13AF1D84475A1361 CRC64;
 Query Match 4.6%; Score 187; DB 1; Length 705;
 Best Local Similarity 21.0%; Pred. No. 0.018;
 Matches 105; Conservative 77; Mismatches 165; Indels 152; Gaps 22;
 QY 6 NAKSSGSSSSSGSGSGSTAGS-----SSPGARRETGHGHNKRGKGLSG 51
 Db 8 NASTTTVIDSKNGSVPKSVKRVKRTVTEDLVTFSSPAA----- 48
 QY 52 TSFFTWNVIALLGWTSVAVVMFDLVYEEV---LGKLG-----LYDA----- 92
 Db 49 -----WLVIALITWSAVAVVMDLVYKNFASASIAKMGSDPLKLVHDAVEETDWTY 103
 QY 93 -----DGDGFDVDDAKVLLG-LKERSTSEPAVPPEAEAPHTPEQVFPV--- 136
 Db 104 GFFSLSDLISSDGEDDEDEDTAKGEIEBPPLKRDHKEKIEKQEKPERKIPTKVV 163
 QY 137 ----EAEQNIIDEAKEQIQSLHEMVAHEHVEGEDLQOEDGPTGEPQOEDDEFLMATDV 192
 Db 164 HKEKEKEKEKVKKEKPE-----KKAHKEKLEKEKEKETKTVTKEKKARTR----- 211
 QY 193 DDRFETLEPEVSHETESHVEETVSQDCNODMEMMSEQENPDSSPEVVEDERLHMDT 252
 Db 212 ----EKIEKTKKE-----VKGVQKEKVKQTVAKAKEVQKTPKPK-----EKSKET 254
 QY 253 DDVTYQYEEQAVY----EPLNEGIEITEVTAPEDNPVEDSQVIVEEVSIFPVEEQ 307
 Db 255 AAVSKQEOXQYAFCRYMIDIFVHGDLPKPGQSPAIPPSPTEQAS--RPTPALPTPEEK 311
 QY 308 E-----VPETNRKTDDBEQAK-----VKKKP-KLINKFDKTKAELDAAE 349
 Db 312 EGEKKAKKKTETTKKAEKEDAKKSEKETDIDMKKEKPGKSPDTKPGTVTTQAT 371
 QY 350 KLRKRGKLEAVNAFKELVRKYPQSPRARYGKACQCEDDLAEKRRSNEVLRAIETTYQEVA 409
 Db 372 K--KDEKEDSKKA-KKPAEEQDK-----GKKQ-----EKKEKHE--EPAKSTKEHA 414
 QY 410 SLPDVPADLLKLSLKRSD 428
 Db 415 APSEKQA---KAKIERKEE 430
 RESULT 8
 CALD_CHICK STANDARD; PRT; 771 AA.
 AC P12957; Q03698; Q90756; Q90761; Q92018; Q99230;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Caldesmon (CDM).
 GN CALD1 OR CAD.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).

RC TISSUE=Gizzard, and Oviduct;
 RX MEDLINE=89340480; PubMed=2760048;
 RA Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
 RT "Cloning and expression of a smooth muscle caldesmon.";
 RL J. Biol. Chem. 264:13873-13879(1989).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
 RC TISSUE=Gizzard;
 RX MEDLINE=90026426; PubMed=2803315;
 RA Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
 RT "Primary structure and functional expression of h-caldesmon complementary DNA.";
 RL Biochem. Biophys. Res. Commun. 164:503-511(1989).
 RN [3]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).
 RC TISSUE=Brain;
 RX MEDLINE=91093148; PubMed=1824698;
 RA Hayashi K., Fujio Y., Kato I., Sobue K.;
 RT "Structural and functional relationships between h- and l-caldesmons.";
 RL J. Biol. Chem. 266:355-361(1991).
 RN [4]
 RP SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).
 RX MEDLINE=94071934; PubMed=8250919;
 RA Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
 RT "Common structural and expressional properties of vertebrate caldesmon genes.";
 RL Biochem. Biophys. Res. Commun. 197:145-153(1993).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).
 RC TISSUE=Gizzard;
 RX MEDLINE=92042686; PubMed=1939602;
 RA Bryan J., Lee R.;
 RT "Sequence of an avian non-muscle caldesmon.";
 RL J. Muscle Res. Cell Motil. 12:372-375(1991).
 RN [6]
 RP SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).
 RC TISSUE=Brain;
 RX MEDLINE=94271210; PubMed=8002994;
 RA Yano H., Hayashi K., Haruna M., Sobue K.;
 RT "Identification of two distinct promoters in the chicken caldesmon gene.";
 RL Biochem. Biophys. Res. Commun. 201:618-626(1994).
 RN [7]
 RP SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Gizzard;
 RX MEDLINE=89273666; PubMed=2730665;
 RA Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
 RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyosin-binding domain in tropomyosin T.";
 RL Biochem. Biophys. Res. Commun. 161:38-45(1989).
 RN [8]
 RP SEQUENCE OF 498-525.
 RX MEDLINE=88293484; PubMed=3401222;
 RA Mornet D., Audemard E., Derancourt J.;
 RT "Identification of a 15 kilodalton actin binding region on gizzard caldesmon probed by chemical cross-linking.";
 RL Biochem. Biophys. Res. Commun. 154:564-571(1988).
 RN [9]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=92041815; PubMed=1939059;
 RA Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
 RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of phosphorylation sites.";
 RL J. Biol. Chem. 266:19971-19975(1991).
 RN [10]
 RP PHOSPHORYLATION OF TYR-27 AND TYR-165.
 RX MEDLINE=20026923; PubMed=10559276;
 RA Wang Z., Danielson A.J., Mathie N.J., McManus M.J.;
 RT "Tyrosine phosphorylation of caldesmon is required for binding to the Shc.Grb2 complex.";
 RL J. Biol. Chem. 274:33807-33813(1999).
 CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the

QY 460 LGVGYLLIGDNDNAKKVYEEV-----LSVTPNDGFA---KVHYGFLKAQN-KIA 505
Db 547 QEEAEERKIREEEKRMRKEIERRAAEAKRQKVPEGVSEEEKPKFCFSPKSSSLKIE 606
QY 506 ESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNK-EAYKWYELGHRG----- 554
Db 607 BRAEFINKSAQK-----SGMKPAHTTAVSVSKIDSRLEQYTSAVGNKAKAPKPAASDL 660
QY 555 -----HFASVWQR-SLYNVNGLKAQPMWTPKET-GYTELVKSLERNWKLIRDEGLAV 604
Db 661 PVPAEGVNRKSMWEKGNVFPSPGCTGP---NKETAGLKVGVSRRINEMWLTTPEG--- 714
QY 605 MDAKAGLEPDEDNLRKGDWS-QFTLMQGRRENENACKGAPKCTLLLEKPEPTTGCCR 662
Db 715 -NKSPA---PKPSDLR-PGDVSGKKNLWEKQSVKEKPASSSKVTAT--GKKSFTNGLRQ 766

RESULT 9

TRDN HUMAN
ID TRDN HUMAN STANDARD; PRT; 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
and its localisation to chromosome 6q22-6q23.";
RL Eur. J. Biochem. 233:258-265(1995).
CC
CC -1- FUNCTION: May be involved in anchoring calsequestrin to the
CC functional sarcoplasmic reticulum and allowing its functional
CC coupling with the ryanodine receptor (By similarity).
CC -1- SUBUNIT: Homooligomer of variable subunit number; disulfide-linked
CC (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; U18985; AAA75315.1; -.
DR PIR; S68191; S68191.
DR Genew; HGNC:12261; TRDN.
DR MIM; 603283; -.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0006936; P:muscle contraction; TAS.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
FT INIT MET 0
FT DOMAIN 1 46
FT TRANSMEM 47 67
FT DOMAIN 68 728
FT DISULFID 269 269
FT DISULFID 690 690
FT CARBOHYD 74 74
FT CARBOHYD 646 646
SQ SEQUENCE 728 AA; 81423 MW; C1C53BBE1B2A0815 CRC64;

Query Match 4.4%; Score 178.5; DB 1; Length 728;

Best Local Similarity 19.2%; Pred. No. 0.049;
Matches 94; Conservative 86; Mismatches 167; Indels 143; Gaps 20;

QY 6 NAKSSGNSSSGSGSGSTAGS-----SSPGARRETKHGHNKRGKGLSG 51
Db 8 NASTTTVIDSKNGSVPRKPGKVLKRTVEDIVTTFFSSPA----- 48
QY 52 TSFTWFMVIALLGWMTSYAVWFDLVYEEV-----LGKLG----- 90
Db 49 -----WLVIALITWSAVAIVMEDLVYKNFSASSLAKIGSDPLKLVRDAMEETDWIY 103
QY 91 -----DADGDGDFVDDAKVLGLKERSTSEPAVPPEAEHPHTEPEQVPV 136
Db 104 GFFSLSDIISSEDEDDGDEDTDKEI---DEPPLRKEIHKDKTEKQEKERKIQT 159
QY 137 EAEPNIEDEAKEQIQSLHEMVHAHVEGEDLQOEDEGPTGEPQOEDEDFLMATDVDDRF 196
Db 160 KVTNKE-KEKGEKVR-KEKPEKKATHKEKEKEKEPETK-----TVAKEQK 205
QY 197 ETLPEVSHETESHYVEETVSQDCNQDEEMMSQENPDSSPVEDERL---HHD 252
Db 206 KAKTAEKSEKTKK--EVKGGQEKVKQTAQKVEVQKTPSKPEKEDKKAASKHEQK 263
QY 253 DQVATQVY-----EQAVYEPLENEGTEITEVTAPEDNPVEDSQIVVEVS 299
Db 264 DQVAFCRMIDIFVHGDLKPGQSPALPPPLPTE-----QASRPTPASPA-----LEEKE 312
QY 300 IFPVEEQEVPETNRK-TDDPEQAK-----YKKKP-KLINFDTIKALDAA--- 348
Db 313 GEKKKAEEKVTSETKKKEKEDIKKSEKETALDVEKKEPGKASETKGTGTVKAAQAANK 372
QY 349 -----EKLRRGKIEAVNAFKELVRKY---PQSPRAYGKAQCEDDLAEKRSNEVL 398
Db 373 DEKKEDSKTKKPAVEVEQPGKQKQEKKEKHEVPAKSPKKEHSVPDKQVAKATERAKEI 432
QY 399 RGAITYQEV 408
Db 433 -GAVSSKKAV 441

RESULT 10

NFM RAT
ID NFM RAT STANDARD; PRT; 845 AA.
AC P12839; Q63370;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NEF3 OR NEFM OR NFM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282618; PubMed=2441012;
RA Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT "Complete amino acid sequence and in vitro expression of rat NF-M,
RT the middle molecular weight neurofilament protein.";
RL J. Neurosci. 7:2590-2599(1987).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Wistar;
RC MEDLINE=92332596; PubMed=1321159;
RX Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT "Schwann cells of the myelin-forming phenotype express neurofilament
RL protein NF-M.";
RL J. Cell Biol. 118:397-410(1992).
RN [3]
RP PHOSPHORYLATION SITES, AND REVISION TO 500.
RX MEDLINE=92165797; PubMed=1537832;
RA Xu Z.-S., Liu W.-S., Willard M.B.;

RT "Identification of six phosphorylation sites in the COOH-terminal
RT tail region of the rat neurofilament protein M.";
RT J. Biol. Chem. 267:4467-4471(1992).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=93346421; PubMed=8344946;
RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
Hart G.W.;
RT "Glycosylation of mammalian neurofilaments. Localization of multiple
RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RT L and M.";
RT J. Biol. Chem. 268:16679-16687(1993).
CC -1- FUNCTION: Neurofilaments usually contain three intermediate
CC filament proteins: L, M, and H which are involved in the
CC maintenance of neuronal caliber.
CC -1- PTM: There are a number of repeats of the tripeptide K-S-P, NFM is
CC phosphorylated on a number of the serines in this motif. It is
CC thought that phosphorylation of NFM results in the formation of
CC interfilament cross bridges that are important in the maintenance
CC of axonal caliber.
CC -1- PTM: Phosphorylation seems to play a major role in the functioning
CC of the larger neurofilament polypeptides (NF-M and NF-H), the
CC levels of phosphorylation being altered developmentally and
CC coincident with a change in the neurofilament function.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18628; AAA41696.1; -;
DR EMBL; Z12152; CAA78136.1; -;
DR PIR; A45669; A45669.
DR GlycoSuiteDB; P12839; -;
DR InterPro; IPR006821; Filament_head.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR Pfam; PF04732; filament head; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurex; Phosphorylation;
KW Glycoprotein.
FT INIT MET 0 0
FT DOMAIN 1 103
FT DOMAIN 103 410
FT DOMAIN 411 844
FT DOMAIN 103 134
FT DOMAIN 135 147
FT DOMAIN 148 246
FT DOMAIN 247 263
FT DOMAIN 264 285
FT DOMAIN 286 289
FT DOMAIN 290 410
FT CARBOHYD 47 47
FT CARBOHYD 430 430
FT MOD_RES 502 502
FT MOD_RES 506 506
FT MOD_RES 536 536
FT MOD_RES 603 603
FT MOD_RES 608 608
FT MOD_RES 666 666
FT MOD_RES 666 666
FT MOD_RES 17 17
FT MOD_RES 21 21
FT MOD_RES 204 204
FT MOD_RES 500 500
SQ SEQUENCE 845 AA; 95660 MW; 316C41655B11197D CRC64;

Query Match 4.4%; Score 176.5; DB 1; Length 845;
Best Local Similarity 23.3%; Pred. No. 0.074;
Matches 90; Conservative 59; Mismatches 137; Indels 101; Gaps 15;
QY 81 EEVLGKLGIDADGDDFDVDDAKVLLGLKERSTSEPAVPPEEAHPHTEPEEQVPEAE 140
Db 453 EEIEETKVEDEKS---EMEDALTYAEELAAAK-----EKEEAEEKEEPEVEKSP 503
QY 141 QNIDBAKEQIQSLHEMVAHEVEGEDQEDGPTGEPQEQEDFLMATDVDRFETLE 200
Db 504 VK-SPEAKEE-----EEGKHEEEEGQ--EEEEEDGVKSDQAEEGGS--E 545
QY 201 PEVSHEETESHVVEETVSQDCNQDMEMNSEQENPDSSPEVVEDRLHHTDVTQVY 260
Db 546 KEGSSEKDE-----GEQEEGETEAEGEGEAKEKEKTEGV 584
QY 261 EEQAVEPEPLENGIE-----ITEVTAPPEDNPVEDSQVIVEVSIFFVEEQQVEP 310
Db 585 EEMAIKEIKVKEPEKAKSPVPKSPVEEVKPKPEAKAGKDEQKEEK-----VEEKEEVA 639
QY 311 PET--NRKTDPEQAK-----VKKKPKLTKFKTKIAELDAEKLK----- 353
Db 640 KESPKKEKVEKEEKKPKDVPDKKKAESPVEKEKAVEMITITKSVKSLKDTKEKKPQQ 699
QY 354 ---RGKIEEAVNAFKELVRKYPQSPRAR-----YGAQCEDDLAEKRSNEVLRAI 402
Db 700 EKVKKEAEEGSGSEEEVGDKSPQESKEDIAINGEVEGKEEEDQTEKSGQEEKEGV 759
QY 403 ETYQEVASLPDYPADLLKLSKRRSDR 429
Db 760 TNGLDVS-----PAE-----EKKGEDR 776

RESULT 11
MY33_HUMAN STANDARD; PRT; 2004 AA.
ID MY33_HUMAN
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MY33 histone acetyltransferase 3 (Runt-related transcription factor
DE binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
DE finger protein 220).
DE GN MY33 OR RUNXBP2 OR ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
Horsman D., Mittleman F., Volinia S., Watmore A.E., Housman D.E.;
RA "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -1- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC translocation t(8;16)(p11;p13) involving MY33 and CREBBP.
CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -1- SIMILARITY: Belongs to the MY33 (SAS/MOZ) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; U47742; AAC50662.1; -.
DR Genew; HGNC:13013; MYST3.
DR MIM; 601408; -.
DR GO; GO:0006323; P:DNA packaging; TAS.
DR InterPro; IPR005818; H1tone_H1/H5.
DR InterPro; IPR002717; MOZ_SAS.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF01853; MOZ_SAS; 1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00526; H15; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
DR PROSITE; PS50016; ZF_PHD_2; 2.
KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1697 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLLOCATION TO FORM MOZ-CBP.
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;

Query Match 4.4%; Score 175; DB 1; Length 2004;
Best Local Similarity 22.7%; Pred. No. 0.26;
Matches 73; Conservative 48; Mismatches 113; Indels 88; Gaps 13;

QY 122 EEAHPTEPEEQVVEAEFQNIIDEAKQIQSLHEMVHAHEVGEDLQOEDGPTGEPOQ 181
Db 1205 QESSETEVEPKEDMP-----LPEERKEE-EMQAEAEAEEGEEDDAASSEVPASPAD 1256
QY 182 EDDEFMATVDVDRFETLEPEVSHETESYHVEETVSQDCNQDMEEMSEQENPDSEP 241
Db 1257 SSN-----SPETETKEPEVEEEKEKRVSEQQRQSEEQELEPEPEEEEDAAFT 1308
QY 242 VVEDERLHHTDD-----VTYQVVEQAVYEPLENE-GI----- 274
Db 1309 AQNDND--HDADDEDDGHLSTKKKELEEQPTREDVKKEPGVQESFLDANMOKSREKIKL 1365
QY 275 -EITEVTAPPEDNPEVDSQVIVEVSIPEVEEQEVPPEPTRKTDDPEQAKVKKKKPKL 333
Db 1366 KEETELDS-EEEQESHDTSVSEQMA-----GSEDDHEDSHTKEELTEL 1409
QY 334 LNKFDKTIKAEIDAAEKLRKRGKIEEAVNAFKELVRKYPOSRRARYGKAQCEDDLAEKRR 393
Db 1410 KEE-EEIPHSELDL-----ETVQAVQSLTQEESSHEGAY--QDCETTLA---- 1451
QY 394 SNEVLRGAIETVQEVASLPDV 415
Db 1452 -----ACQTLQSYTQADEDP 1466

```

RESULT 12

```

YNT1_YEAST STANDARD; PRT; 1240 AA.
ID YNT1_YEAST
AC P53935;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
GN YNL091W OR N2231.
OS Saccharomyces cerevisiae (Baker's yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96367601; PubMed=8771715;
RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
RT "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
RT chromosome XIV that includes the YPT53, trnAleu and gsr m2 genes and
RT four new open reading frames.";
RL Yeast 12:599-608(1996).
CC -!- SIMILARITY: TO S.POMBE SPAC29E6.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X85811; CAA59826.1; -.
CC EMBL; Z71367; CAA95967.1; -.
CC PIR; S52734; S52734.
CC Germonline; 143097; -.
CC SGD; S0005035; YNL091W.
CC GO; GO:0009651; P:salinity response; IMP.
KW Hypothetical protein.
FT DOMAIN 756 761 POLY-GLU.
SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64;

```

```

Query Match 4.3%; Score 172.5; DB 1; Length 1240;
Best Local Similarity 23.9%; Pred. No. 0.19;
Matches 103; Conservative 66; Mismatches 175; Indels 87; Gaps 21;

```

```

QY 144 EDBAKEQIQSLHEMVHAHEVE-GEILQOEDGPTGEPOQEDDEFMATVDVDRFETLEPE 202
Db 460 EDEDEEDYDD-----YSEYAEDESHVESEYEGIEAVEKEPHE-----KSGIRETILHS 508
QY 203 VSH-----ETESYHVEETVSQDCNQDMEEMSEQENP-----DSSEP 241
Db 509 YDHDHKRQNHPHHHYSTHSED-ELSEEYISDIELPHDPKHFRDDDLIDGDEDEP 567
QY 242 VVEDERLHHTDDVTYQ--VYEQAVYEPLNEGIEITE-----VTAPPEDNPEVDSQV- 293
Db 568 EEDENEGDDEED-TYDSGLDETDRLEEGKLIQIATKLIQSRIMASYHEKQADNNRLK 626
QY 294 IVEEVSIFPEVEEQEVPPEPTRKTDDPEQAKVKKKKPKLTKKFDK-----TIKAEI 345
Db 627 LLQE-----LEEEKRKREKEKEKQKREKEKEKRLQQLAKEEERKRREBEKERLKKEL 681
QY 346 DAAEKLRK---RGKIEEAVNAFKELVRKYPOSRRARYGKAQCEDDLAEKRR--SNEVLRG 400
Db 682 EBERMRREARQKVEEA-----KKKDEBRKRRLLEQQRREEMQEKQRKQKEELKKR 734
QY 401 AIETVQEVASLPDVPA DLKLSLKRSRDRQOFLGHMRGSLTLQL--VQLFPNDTSLK- 457
Db 735 REEEKRRIHQKRLQEKLOKE-KHEERQRLIAE--DALRKQKLNEQTSANILSAKP 790
QY 458 -NDLGVGYLLIGDNDNAKKVVEEIVSVTPNDGFAKVHYGLIKAKNKIAESIPYLKEGLE 516
Db 791 FTENGVGNPVSSQSHPMNTNYQEDNSCSINDE-----ILKNVNSVAASKPVSPGTGFN 842
QY 517 SGD---PGTDD 524
Db 843 VHDLLPSTNN 853

```

RESULT 13

```

YHU6_YEAST STANDARD; PRT; 465 AA.
ID YHU6_YEAST
AC P38845;

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 51.1 kDa protein in DCD1-MRPL6 intergenic region.
GN YHR146W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Giesel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VII.";
RL Science 265:2077-2082(1994).
CC -!- SIMILARITY: TO YEAST YNL173C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U10397; AAB68982.1; -.
DR PIR; S46759; S46759.
DR GerMOnline; 139464; -.
DR SGD; S0001189; CRP1.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003677; F:DNA binding; IDA.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 5115 MW; 30880758F37991C7 CRC64;

Query Match 4.3%; Score 172; DB 1; Length 465;
Best Local Similarity 22.6%; Pred. No. 0.06;
Matches 90; Conservative 52; Mismatches 132; Indels 124; Gaps 16;

QY 76 DLVDYBEVLGKLGITYDADSGDFDVKAKVLG----- 108
DB 92 DLVETQEVAGASRIPEAG-----LIGKPPRSAGPSTSNRKNKRNKRRSK 141
QY 109 LKERSTSEPAVPPEAEPTPE-----EQVPEAEPONIIDEAKEQIQS 153
DB 142 LKKKSTKNKKSNESLDNNEEDGVTGTTEDVTGTSREETPL-AEPTNVSKAEPGFHI 200
QY 154 LHEMVHAHEVGEDLQOEDGPTGER-----QOEDDEFMATVDVDRFETLEPEVSHE 206
DB 201 LPTD-----QSADITGNGIIGSPGVLVPNGEIKEFTEIRVDVAR-----E 243
QY 207 ETEHSYHVEETVSQDCNQDMEEMSEQEN--PDSSEPVVEDERLHHTDDVTYQYEEQA 264
DB 244 LNERLNKKEEVPPEVAGPIVESSVTEKSPALPADDPVETKEVAHNVQELTPQV--EA 300
QY 265 VYEPLENEGIEITEVTAPEDNPVEDSQVIVEEVS-IPVVEEQQEVPEPTNRKTDPEQK 323
DB 301 V-TPLINE-----PEPLPTPEAQISIPSSKVEPEVGSIQ----- 334
QY 324 AKVKKKKPKLINFDKTIKAELDAEKLRRKGKIEAVNAFKELVRKYPQSP-----RAR 378
DB 335 -----SKLVEKREST-EGVLDSKKVENKAKDEEVFTLDPIVNAKPKLPLTDEQTAE 386
QY 379 YGKAOCEDDLAEKRRSNEVLGALETYQEVASLPDVPA 416
DB 387 GRKSPAVSESEKKEKKQOE--KGSKEVKRSETSEKKEKPS 422

RESULT 14
YCF2_OENV1
ID YCF2_OENV1 STANDARD; PRT; 630 AA.
AC P31569;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycf2 (ORF 2280) (Fragment).
GN YCF2.
OS Oenothera villaricae.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Myrtales; Onagraceae; Oenothera.
OX NCBI_TaxID=3941;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93169690; PubMed=8435856;
RA Nimzyk R., Shoendorf T., Hachtel W.;
RT "In-frame length mutations associated with short tandem repeats are
RT located in unassigned open reading frames of Oenothera chloroplast
RT DNA.";
RL Curr. Genet. 23:265-270(1993).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64615; CAA45896.1; -.
DR PIR; S29796; S29796.
KW Chloroplast; Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 630 AA; 72781 MW; 6AEFFFD7C75B0BAA CRC64;

Query Match 4.3%; Score 172; DB 1; Length 630;
Best Local Similarity 20.9%; Pred. No. 0.087;
Matches 107; Conservative 82; Mismatches 168; Indels 154; Gaps 22;

QY 52 TSFFTF-----MVIALGVWTSVAVVWFDLV-----DYEEVLGKLGITYDADGDD 97
DB 100 SLYKWFYFELGTSMKKLTILLYLTCSAGSIAQDLSPGPDEQNLITSYGLVENDSDLV 159
QY 98 FDVDDAKVLGLKERSTSEPAVPPEAEPTPEBQVPEAEPONIIDE--AKEQIQSL 154
DB 160 HGLSD--IVHGLILEGALVGSSFTVEEVEGTEEEVEGTEEEVEGTEEEVEGTE 217
QY 155 LHEMVHA--EHVEG-----EDLQOEDGPTGERQOEDDEFMATVDVDRFETLEPEV--SH 205
DB 218 EDEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTE--DEVEGTEEEVEGTE 276
QY 206 ETEHSYHVEETVSQDCNQDMEEMSEQENPDSSEPVV--EDERLHHTDDVT----- 256
DB 277 EEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEKDSQFDNDRVTLRLRPK 336
QY 257 -----YQVE-----EQAVYEP--LENEGIEIT----- 277
DB 337 PRNPLDIQRLIYQHOKYSESELEDDDDDDVFAFQKMLDLFSELVWSPRIWHPWDFLLD 396
QY 278 -EVTAPEDNPVEDSQV--IVEEVSIFPVEE-----QOEVPEPTNRK 316
DB 397 CEAEIPAEIPEEDELDPEDALETEVAWGVVEEGEADDEEDVLLEAQOEDELLEEDDEE 456
QY 317 TDDPEQAKVKKKKPK---LINFDKTIKAELDAEKL----- 351
DB 457 LDEEDELDEEEBEPKKEEDELHEEEHEEEHEEEHEEDELQENDSEFFRVKPIIPRRWIF 516

QY 352 RRGKI-----EEAVNAFKELV-----RKYPSPRARYGKAQCEDD----- 387
Db 517 RKKDVFVLSYPEEATEISKELLRLNPKTKRDAPKRPQRQWTKKKQDKHYELLDRQ 576
QY 388 --LAEK--RSNEVLRG--AIETYQEVASL 411
Db 577 RMLITRSLSKSNGFPRSNTPSESYQYLSNL 607

RESULT 15
LAMB2_MOUSE STANDARD; PRT; 592 AA.
ID LAMB2_MOUSE
AC P21619;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lamin B2.
GN LMNB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91106216; PubMed=2102682;
RA Hoefer T.H., Zatloukal K., Waizenegger I., Krohne G.;
RT "Characterization of a second highly conserved B-type lamin present
in cells previously thought to contain only a single B-type lamin.";
RL Chromosoma 99:379-390(1990).
RN [2]
RP ERRATUM.
RX MEDLINE=91339548; PubMed=2102440;
RA Hoefer T.H., Zatloukal K., Waizenegger I., Krohne G.;
RL Chromosoma 100:67-69(1990).
CC -1- FUNCTION: Lamins are components of the nuclear lamina, a fibrous
layer on the nucleoplasmic side of the inner nuclear membrane,
which is thought to provide a framework for the nuclear envelope
and may also interact with chromatin.
CC -1- SUBCELLULAR LOCATION: Nucleoplasmic side of the inner nuclear
membrane.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=B2;
CC IsoId=P21619-1; Sequence=Displayed;
CC Name=B3;
CC IsoId=P48680-1; Sequence=External;
CC Note=No experimental confirmation available;
CC -1- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS
PHOSPHORYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF
THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY
PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.
CC -1- MISCELLANEOUS: The structural integrity of the lamina is strictly
controlled by the cell cycle, as seen by the disintegration and
formation of the nuclear envelope in prophase and telophase,
respectively.
CC -1- SIMILARITY: Belongs to the intermediate filament family. THIS IS A
B TYPE LAMIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54098; CAA38032.1; -.
DR PIR; B48315; B48315.
DR MGD; MGI:96796; Lmb2.
DR GO; GO:0005638; C:Lamin filament; IDA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR001322; IF_tail_C.
DR Pfam; PF00038; filament; 1.

DR Pfam; PF00932; IF_tail; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Nuclear protein; Lipoprotein;
KW Preylation; Phosphorylation; Alternative splicing.
FT DOMAIN 1 26 HEAD.
FT DOMAIN 27 378 ROD.
FT DOMAIN 379 592 TAIL.
FT DOMAIN 27 61 COIL 1A.
FT DOMAIN 62 73 COIL 1B.
FT DOMAIN 74 207 LINKER 1.
FT DOMAIN 208 234 LINKER 2.
FT DOMAIN 235 378 COIL 2.
FT DOMAIN 414 419 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 561 574 ASP/GLU-RICH
FT LIPID 589 589 S-farnesyl cysteine (By similarity).
FT SEQUENCE 592 AA; 67029 MW; 7D5AC51BC9A5041E CRC64;
SQ

Query Match 4.2%; Score 169.5; DB 1; Length 592;
Best Local Similarity 20.6%; Pred. No. 0.11;
Matches 129; Conservative 113; Mismatches 244; Indels 141; Gaps 29;

QY 78 VDYEYVTLGKIGYDADGDDGFVDYDAKVLG-LKERSTSEPAVPPEAEPTPEEQVP 135
Db 62 VTTREVSGIKTLYES-----ELADARRVLDETARERARLQIEIGKVQALBEARKSAKK 115
QY 136 VEAPQNTDEAKEQIQSLH---EMVHA---EHVEGEDLQ-----QEDGPTGHPQ 180
Db 116 REGELTVAQGRVKD-LESLFHRSEAEALATALSDNEGLETEVAELRAQLAKAEDGHAVAKK 174
QY 181 QEDDEFMATVDVDRFETLEPEVSHETESHYHVEETVSQDCNQDMEEMSEQENPD-SS 239
Db 175 QLEKETLMRVDLNRCQSLQEEELAFSKSVFEEVRETRRHERRLIVEVDSRRQEQYDFKM 234
QY 240 EPVVEDERLHDDTDVTVQVVEQAVVEPLENEGIEITEVTAPBEDNPVEDSQVIVEVS 299
Db 235 AQALEDLRSQHDEQVRLYRVELEQTYQAKLDNAKL----- 269
QY 300 IFPVEEQEVPPEPTRKTDDPEQAKAVKKKPKLANKFDKTIKALDAEKLRRGKIEE 359
Db 270 ---LSDQNDKAHAARE-ELKEARMVESLSYQLG-----LQKQASAAE--NHIELEE 318
QY 360 AVNAFKELVRKYPSPRARYGKAQCEDDLAEKRRSNEVLRGAIEITYQEVASLP---DYPA 416
Db 319 ALRGERDKFRKMLDAK-----EQEMTEVR--DRMOQLAEYQELLDIKALDMEI 366
QY 417 DLLKLSKRSDRQQFLGHMRSLLTLQRLVOLFPNDTSLKNDLGVCYLLIGDNNAKV 476
Db 367 SAYRKLIEGEEERLK-LSPSPSSRITISRA-----TSSSSSSSGVG-MSVGQRGGRRR 417
QY 477 YEEVLSYTPNDGFAKVHYGLIKAKNKIAESIPYLKEGIESGDPGTDGRF--YFHLGDA 534
Db 418 LEDT-SGSPSRA-SRVSSG-----SRLAQTVATGVNVIDEVP--EGRFVRLKSSDK 466
QY 535 MORVGNKE-----AYKN---YELGHRGHFASVWQRSLYNVNG-----LKAQP 574
Db 467 DQSLGNWRIRKQVLLEGEDIAKFTPKYVL--RAGQTVTWAAAGAGATHSPSTLVWKSQT 524
QY 575 WWTPEKTEYTELKSLERNWKLIRDEGLAVMDKAKGLFLEDENLREKGDWSQFTLMQGG 634
Db 525 NMGPGESRITALVSADGEVAVKAKHSSVQGRENG-----EEEEEEAEFGEEDLFHQ- 578
QY 635 RRNENACKGAPKCTTLLEKFPETTGCR 661
Db 579 -----QGDPT-----TSRGCR 590

Search completed: May 19, 2004, 15:58:21
Job time : 22.2142 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:54:24 ; Search time 73.0139 Seconds
(without alignments)
3275.577 Million cell updates/sec

Title: US-09-903-199-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSSGSGS.....IVDVHPELTPQQRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4022	100.0	758	4	Q9Y4J0	Q9Y4J0 homo sapien
2	3209.5	79.8	739	11	Q9EPA6	Q9EPA6 mus musculu
3	3208.5	79.8	741	11	Q8BSY0	Q8BSY0 mus musculu
4	3175.5	79.0	725	11	Q8BQK0	Q8BQK0 mus musculu
5	2950	73.3	658	11	Q8CBM2	Q8CBM2 mus musculu
6	2919.5	72.6	689	11	Q9EQ66	Q9EQ66 mus musculu
7	1636	40.7	313	4	Q9H2C4	Q9H2C4 homo sapien
8	1457.5	36.2	299	4	Q9H291	Q9H291 homo sapien
9	1376.5	34.2	270	4	Q8TB28	Q8TB28 homo sapien
10	1066.5	26.5	785	5	Q9GQ82	Q9GQ82 drosophila
11	998.5	24.8	308	11	Q9EQ65	Q9EQ65 mus musculu
12	972.5	24.2	292	11	Q9IWG6	Q9IWG6 mus musculu
13	890	22.1	270	11	Q920F9	Q920F9 mus musculu
14	798	19.8	872	5	Q93178	Q93178 caenorhabdi
15	779.5	19.4	259	11	Q920F8	Q920F8 mus musculu
16	708.5	17.6	258	11	Q9EQ67	Q9EQ67 mus musculu

17	691.5	17.2	215	11	Q920F7	Q920F7 mus musculu
18	550.5	13.7	133	11	Q9EQ63	Q9EQ63 mus musculu
19	533.5	13.3	147	11	Q9EQ62	Q9EQ62 mus musculu
20	516	12.8	245	11	Q8CH79	Q8CH79 mus musculu
21	435.5	10.8	210	4	Q9NR11	Q9NR11 homo sapien
22	420.5	10.5	210	6	Q28264	Q28264 canis fami
23	418	10.4	225	4	Q9NR10	Q9NR10 homo sapien
24	407.5	10.1	207	11	Q9EQ64	Q9EQ64 mus musculu
25	407.5	10.1	207	11	Q9CR06	Q9CR06 mus musculu
26	407.5	10.1	212	11	Q9D7J8	Q9D7J8 mus musculu
27	390.5	9.7	304	2	Q53792	Q53792 streptomyc
28	276	6.9	343	11	Q80VP9	Q80VP9 mus musculu
29	275	6.8	250	16	Q93H17	Q93H17 streptomyc
30	274.5	6.8	343	4	Q9UHS3	Q9UHS3 homo sapien
31	263.5	6.6	186	4	Q9NSN3	Q9NSN3 homo sapien
32	257	6.4	176	6	Q9NIE7	Q9NIE7 oxyctolagus
33	251	6.2	176	11	Q9CUZ2	Q9CUZ2 mus musculu
34	250	6.2	49	11	Q9EQ69	Q9EQ69 mus musculu
35	217.5	5.4	230	4	Q8N4H3	Q8N4H3 homo sapien
36	217.5	5.4	250	4	Q8N316	Q8N316 homo sapien
37	217.5	5.4	280	4	Q8IW63	Q8IW63 homo sapien
38	217.5	5.4	347	4	Q96H00	Q96H00 homo sapien
39	216.5	5.4	1110	13	Q91255	Q91255 petromyzon
40	207.5	5.2	1236	5	Q9GTX2	Q9GTX2 plasmodium
41	206	5.1	17352	5	Q95YM2	Q95YM2 procambarus
42	204.5	5.1	3111	5	Q9VH10	Q9VH10 drosophila
43	204	5.1	976	12	Q9DUN0	Q9DUN0 kaposi's sa
44	203	5.0	1233	5	Q8IJ56	Q8IJ56 plasmodium
45	199	4.9	1271	5	Q25860	Q25860 plasmodium

ALIGNMENTS

RESULT 1

Q9Y4J0 PRELIMINARY; PRT; 758 AA.
ID Q9Y4J0
AC Q9Y4J0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Aspartyl(asparaginy)l-beta-hydroxylase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96420598; PubMed=8823296;
RA Lavalaisiere L., Jia S., Nishiyama M., de la Monte S., Stern A.M.,
RA Wands J.R., Friedman P.A.;
RT "Overexpression of human aspartyl(asparaginy)l-beta-hydroxylase in
RT hepatocellular carcinoma and cholangiocarcinoma.";
RL J. Clin. Invest. 98:1313-1323(1996).
DR EMBL; S83325; AAB50779.1; -;
DR InterPro; IPR007943; Asp-B-hydro N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 758 AA; 85890 MW; 4AF6F0AB4500AF0C CRC64;

Query Match 100.0%; Score 4022; DB 4; Length 758;
Best Local Similarity 100.0%; Pred. No. 9.1e-232;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGNSSSSSGSGSTSAGSSSPGARRTKHGKNGRKGLSGTSFTFMV 60
DB 1 MAQRKNAKSSGNSSSSSGSGSTSAGSSSPGARRTKHGKNGRKGLSGTSFTFMV 60
QY 61 IALLGWTSAVAVVDFLDVYEVLGKLGITDADGDGDFVDVDAKVLGLKXSTSEPAVP 120

Db 61 IALLGVTSVAVWFDLVYBEVLGKLGITYDADGDGDFDVKAVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEPPONIEDEAKEQIQSLHMHWAHEHVEGEDLOQEDGPTGEPO 180
Db 121 PEEAEPHTEPEEQVPVEAEPPONIEDEAKEQIQSLHMHWAHEHVEGEDLOQEDGPTGEPO 180
QY 181 QEDDEFIMATDVDDRFETLEPEVSHTEHSHYVEETVSQDCNODMEEMMSQENPDSSSE 240
Db 181 QEDDEFIMATDVDDRFETLEPEVSHTEHSHYVEETVSQDCNODMEEMMSQENPDSSSE 240
QY 241 PVVEDERLHNDTDDVTYQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
Db 241 PVVEDERLHNDTDDVTYQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDSQVIVEEVS 300
QY 301 FVVEEQEVPPEPNRKTDDPEQAKAVKKKKPKLINKFDKTIKALDAEKLRRGKIEEA 360
Db 301 FVVEEQEVPPEPNRKTDDPEQAKAVKKKKPKLINKFDKTIKALDAEKLRRGKIEEA 360
QY 361 VNAFKELVRKYPQSPRARGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVPADLLK 420
Db 361 VNAFKELVRKYPQSPRARGKAQCEDDLAEKRSNEVLRGAIETTYQEVASLPDVPADLLK 420
QY 421 LSLKRRSDROQFLGHRGSLTLQRLVOLFPNDTSLKNDLGVGYLLIGDNDNAKVVEEV 480
Db 421 LSLKRRSDROQFLGHRGSLTLQRLVOLFPNDTSLKNDLGVGYLLIGDNDNAKVVEEV 480
QY 481 LSVTENDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
Db 481 LSVTENDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHLGDAMQVRGN 540
QY 541 KEAYKMYELGHRGFASVWQSRSLYVNGLKAQPMWTPKETGYTELKSLERNMKLIRDE 600
Db 541 KEAYKMYELGHRGFASVWQSRSLYVNGLKAQPMWTPKETGYTELKSLERNMKLIRDE 600
QY 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTLLLEKFPETTG 660
Db 601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTLLLEKFPETTG 660
QY 661 RRGQIKYSIMHPGTHVWHTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEKGVLIFD 720
Db 661 RRGQIKYSIMHPGTHVWHTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTWEKGVLIFD 720
QY 721 DSFEHEVWQDASSFRLLFIVDVWHPBELTPQRRSLPAI 758
Db 721 DSFEHEVWQDASSFRLLFIVDVWHPBELTPQRRSLPAI 758
RESULT 2
Q9EP6 PRELIMINARY; PRT; 739 AA.
AC Q9EP6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase 6.6 kb transcript (Aspartyl beta-hydroxylase
DE 4.5 kb transcript).
GN 231005F16RIK OR ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction.";
RL J. Biol. Chem. 275:39543-39554 (2000).
DR EMBL; AF289487; AAG40809.1; -.

DR EMBL; AF289486; AAG40808.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 739 AA; 82841 MW; 4DF9F642512CA4EB CRC64;
Query Match 79.8%; Score 3209.5; DB 11; Length 739;
Best Local Similarity 80.7%; Pred. No. 2.8e-183;
Matches 619; Conservative 41; Mismatches 70; Indels 37; Gaps 8;
QY 1 MAQRKNAK-SSGSSSSSGSGSS-----TSAGSSSPGARRETKHGKNGKGLSG 51
Db 1 MAPRKNAKGGGSSSSSGSGSGSGSPSTGSSSSSPGARREAKHGKNGRGGISG 60
QY 52 TSFTFMVIALIGVWTSVAVWFDLVYBEVLGKLGITYDADGDGDFDVKAVLLGLKE 111
Db 61 GSFTFMVIALIGVWTSVAVWFDLVYBEVLGKLGITYDADGDGDFDVKAVLLGLKE 120
QY 112 RSTSEPAVPEEAEPHTEPEEQVPVEAEPPONIEDEAKEQIQSLHMHWAHEHVEGEDLOQ 171
Db 121 RSPSERTPFP-EAETHALEEQAPEGADIQNVEDEVKEQIQSLHMHWAHEHVEGEDLOQ 174
QY 172 EDGPTGEPOQEDDEFIMATDVDDRFETLEPEVSHTEHSHYVEETVSQDCNODMEEMMS 231
Db 175 ADGLAGEPOPEVEDFLVTDDSDRFEDLEPGTVHEIEDTYHVEDTASQNHNDMEEMTN 234
QY 232 EQENPDSSPEVVEDERLHNDTDDVTYQVYEEQAVYEPLNEGIEITEVTAPPEDNPVEDS 291
Db 235 EQENPDSEAVTDAVLLPHAEVVRQDYDE-PVYEPESEHEGEIS-----DNTIDDS 286
QY 292 QVIVEEVSIPVEEQEVPPEPNRKTDDPEQAKAVKKKKPKLINKFDKTIKALDAEKL 351
Db 287 SIISEINVASVEEQDTPP-----VKKKKPKLINKFDKTIKALDAEKL 332
QY 352 RKRKIEEAVNAFKELVRKYPQSPRARGKAQCEDDLAEKRSNEVLRGAIETTYQEVASL 411
Db 333 RKRKIEEAVNAFELVRKYPQSPRARGKAQCEDDLAEKRSNEVLRAIETTYQEAADL 392
QY 412 PDVPADLLKLSLKRSDROQFLGHRGSLTLQRLVOLFPNDTSLKNDLGVGYLLIGDND 471
Db 393 PDAPTDLVKLSLKRSEROQFLGHRGSLTLQRLVOLFPNDTLLKNDLGVGYLLIGDND 452
QY 472 NAKKVEEVSIVTPNDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHL 531
Db 453 SAKKVEEVLNVTTPNDGFAKVHYGFLKAQNKIAESIPYLKEGIESGDPGTDGGRFYFHL 512
QY 532 GDAMQVRGNKAYKMYELGHRGFASVWQSRSLYVNGLKAQPMWTPKETGYTELKSL 591
Db 513 GDAMQVRGNKAYKMYELGHRGFASVWQSRSLYVNGLKAQPMWTPRETGYTELKSL 572
QY 592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTLL 651
Db 573 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTLL 632
QY 652 EKPEPTTGCRGQIKYSIMHPGTHVWHTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTW 711
Db 633 EKPEPTTGCRGQIKYSIMHPGTHVWHTGPTNCRRLRMHLGLVIPKEGCKIRCANETRTW 692
QY 712 EBGKVLIFDSDFEHEVWQDASSFRLLFIVDVWHPBELTPQRRSLPAI 758
Db 693 EBGKVLIFDSDFEHEVWQDASSFRLLFIVDVWHPBELTPQRRSLPAI 739
RESULT 3
Q8BSY0 PRELIMINARY; PRT; 741 AA.
AC Q8BSY0;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Uterus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK030293; BAC26882.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 741 AA; 83042 MW; 0660A6A5E34418C8 CRC64;

Query Match 79.8%; Score 3208.5; DB 11; Length 741;
Best Local Similarity 80.6%; Pred. No. 3.2e-183;
Matches 619; Conservative 41; Mismatches 71; Indels 37; Gaps 8;

QY 1 MAQRKNAK-SSGNSSSSSGSGSGS-----TSAGSSSPGARRETKHGHNKRGKGLSG 51
DB 1 MAPRKNAKGGGSSSSSGSGSGSGSPSTGSSSSSPGARREAKHGHNKRGGRGTSIG 60
QY 52 TSFTWMVIALLGWTSVAVWFVLDVYEVLGKLGTYDADGDGDFDVDADAKVLLGLKE 111
DB 61 GSFTWMVIALLGWTSVAVWFVLDVYEVLGKLGTYDADGDGDFDVDADAKVLLGLKE 120
QY 112 RSTSEPAVPP-EEAEPHTEPEEQVVEAEPQNIIDEAKEQIOSLLHEMVHAHEVEGDLQ 170
DB 121 RSPSERTFPPEEAETHAELEEQAPEGADIQNVEDEVKEQIOSLLQESVHTDH---DL- 175
QY 171 QEDGPTGEPOQEDDEFLMATVDVDRFETLPEVSHETEHSYHVEETVSQDCNQDEEMM 230
DB 176 EADGLAGEPQPEVEDFLTITDSDRFEDLEPGTVHEEIEDTYHVEDTASQNHNDMEEMT 235
QY 231 SEQENPDSSSEPVVEDERLHHTDVTYQVVEEQAVYEPLNEGIEITEVTAPEDNPVED 290
DB 236 NEQNSDPSEAVTDAVGLPHAAEVRAQDYDE-PVYEPSEHEGVAIS-----DNTIDD 287
QY 291 SQVIVEEVSIFPVEEQQEVPEPTNRKTDDEPQAKVKKKKPKLINKFDKTIKAEIDAAEK 350
DB 288 SSITSEEINVASVEEQDTPP-----VKKKKPKLINKFDKTIKAEIDAAEK 333
QY 351 LKRGKIEEAVNAFKELVKYPQSPRARYGKAQCEDDLAEKRKSNEVLGAIETEQEVAS 410
DB 334 LKRGKIEEAVNAFEEIVRKYPQSPRARYGKAQCEDDLAEKRKSNEVLRAIETEQEAAD 393
QY 411 LPDVPADLLKLSLKRSDRQQLGHMGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDN 470
DB 394 LPDAPTDLVKLSLKRSEERQQLGHMGSLLTLQRLVQLFPSPDTTLKNDLGVGYLLIGDN 453
QY 471 DNAKKVVEEVLSTPNDFPAKVHYGFLKAQNKIAESIPYKKGIESGDPGTIDGRFYFH 530
DB 454 DSAKKVVEEVLNTPNDFPAKVHYGFLKAQNKISESIPYKKGIESGDPGTIDGRFYFH 513
QY 531 LGDAMQVRVGNKAYKYWELGHKRGHPSAVWQRSLYNVNGLKAQPMWTPKETGYTELKVSJ 590
DB 514 LGDAMQVRVGNKAYKYWELGHKRGHPSAVWQRSLYNVNGLKAQPMWTPRETGYTELKVSJ 573
QY 591 ERNWKLIRDEGLAVMDKAGLFLPEDENLREKGDWSQFTLWQGRNENACKGAPKTCCTL 650

DB 574 ERNWKLIRDEGLAVMDKAGLFLPEDENLREKGDWSQFTLWQGRNENACKGAPKTCAL 633
QY 651 LEKFPETGCRGQIKYSIMHPGTHVPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRT 710
DB 634 LEKFSSETGCRGQIKYSIMHPGTHVPHGTPTNCRRLRMHLGLVIPKEGCKIRCANETRT 693
QY 711 WEEGKVLIFDPSFEHEWQDASSERLIFVDVWHPPELTPOQRSLPAI 758
DB 694 WEEGKVLIFDPSFEHEWQDASSERLIFVDVWHPPELTPOQRSLPAI 741

RESULT 4

Q8BQKO PRELIMINARY; PRT; 725 AA.
ID Q8BQKO
AC Q8BQKO;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK049506; BAC33783.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 725 AA; 81492 MW; F3CE979F8FC9C3D5 CRC64;

Query Match 79.0%; Score 3175.5; DB 11; Length 725;
Best Local Similarity 79.8%; Pred. No. 2.9e-181;
Matches 613; Conservative 41; Mismatches 61; Indels 53; Gaps 9;

QY 1 MAQRKNAK-SSGNSSSSSGSGSGS-----TSAGSSSPGARRETKHGHNKRGKGLSG 51
DB 1 MAPRKNAKGGGSSSSSGSGSGSGSPSTGSSSSSPGARREAKHGHNKRGGRGTSIG 60
QY 52 TSFTWMVIALLGWTSVAVWFVLDVYEVLGKLGTYDADGDGDFDVDADAKVLLGLKE 111
DB 61 GSFTWMVIALLGWTSVAVWFVLDVYEVLGKLGTYDADGDGDFDVDADAKVLLGLKE 120
QY 112 RSTSEPAVPP-EEAEPHTEPEEQVVEAEPQNIIDEAKEQIOSLLHEMVHAHEVEGDLQ 170
DB 121 RSPSERTFPPEEAETHAELEEQAPEGADIQNVEDEVKEQIOSLLQESVHTDH---DL- 175
QY 171 QEDGPTGEPOQEDDEFLMATVDVDRFETLPEVSHETEHSYHVEETVSQDCNQDEEMM 230
DB 176 EADGLAGEPQPEVEDFLTITDSDRFEDLEPGTVHEEIEDTYHVEDTASQNHNDMEEMT 235
QY 231 SEQENPDSSSEPVVEDERLHHTDVTYQVVEEQAVYEPLNEGIEITEVTAPEDNPVED 290
DB 236 NEQNSDPSEAVTDAVGLPHAAEVRAQDYDE-PVYEPSEHEGVAIS-----DNTIDD 271
QY 291 SQVIVEEVSIFPVEEQQEVPEPTNRKTDDEPQAKVKKKKPKLINKFDKTIKAEIDAAEK 350
DB 272 SSITSEEINVASVEEQDTPP-----VKKKKPKLINKFDKTIKAEIDAAEK 317
QY 351 LKRGKIEEAVNAFKELVKYPQSPRARYGKAQCEDDLAEKRKSNEVLGAIETEQEVAS 410

```

Db      318 LKRGKIEEAVNAFEELVRKYPQSPRARYGKAQCEDDDLAEKQSRNEVLRAIETTYQEAAD 377
QY      411 LPDVPADLLKLSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDN 470
Db      378 LPDAPIDLVLKLSLKRSEKQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDN 437
QY      471 DNAKYVEEVLSTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFH 530
Db      438 DSAKYVEEVLNTPNDGFAKVHYGFIKAQNKISESIPYLKEGIESGDPGTDGRFYFH 497
QY      531 LGDAMQVRGNKEAYKMYELGHRGFASVWQSLYNVNGLKAQPMWTPRETGYTELVKSL 590
Db      498 LGDAMQVRGNKEAYKMYELGHRGFASVWQSLYNVNGLKAQPMWTPRETGYTELVKSL 557
QY      591 ERNWKILRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKTCITL 650
Db      558 ERNWKILRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKTCAL 617
QY      651 LEKPEETTCRRGQIKYSIMHPGTHWPHGTPTNCRRLMHLGLVIPKEGCKIRCANETRT 710
Db      618 LEKSETTCRRGQIKYSIMHPGTHWPHGTPTNCRRLMHLGLVIPKEGCKIRCANETRT 677
QY      711 WEEGKVLIFDDSFHEVWQDASSFRLIFIVDVWHPBELTPQRRSLPAI 758
Db      678 WEEGKVLIFDDSFHEVWQDASSFRLIFIVDVWHPBELTPQRRSLPAI 725

```

RESULT 5

```

Q8CBM2 PRELIMINARY; PRT; 658 AA.
ID Q8CBM2;
AC Q8CBM2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL, AK035735; BAC29171.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro. N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 658 AA; 75127 MW; 2BA0A5C8E06801C8 CRC64;

```

```

Query Match 73.3%; Score 2950; DB 11; Length 658;
Best Local Similarity 80.7%; Pred. No. 7.3e-168;
Matches 566; Conservative 38; Mismatches 53; Indels 44; Gaps 7;

```

```

QY      59 MVTALLGWTSTVAVWFDLVYEEVLGKLGITYDADGDGDFVDDAKVLLGLKERSTSEPA 118
Db      1 MVTALLGWTSTVAVWFDLVYEEVLGKLGITYDADGDGDFVDDAKVLLGLKERSTSEPA 60
QY      119 VPP-EEAEPHTEPEEQVPEAEFQNIIDEAKQIQSLHENVHAHVGEEDLQEDGPTG 177
Db      61 FPPPEEAETHAELEQAEQEGADIQNVEDEVEKEQIQSLQESVHTDH----DL-EADGLAG 115

```

```

QY      178 EPQOEDEFMATVDVDRFETLEPEVSHETESHVHEETVSOQCNQDMSEEMSEQENPD 237
Db      116 EPQPEVEDFLTITDSDDRFEDLEPGTVHEIEDTYHVEDTASQNHPNDMSEEMTNEQEN-- 173
QY      238 SSEPVEDERLHDTDVTQVYEEQAVEPLENEGIEITEVTAPBEDNPVEDSQVYVEE 297
Db      174 -----SEVHRQDYDE-PYEPSEHEGVAIS-----DNTIDSSIISEE 211
QY      298 VSIFPVEEQEVEPEETNRKTDDEQAKAYKKKKPKLINKFDKTIKAELDAEKLARKRGI 357
Db      212 INVASVEEQQDTPP-----YKKKKPKLINKFDKTIKAELDAEKLARKRGI 257
QY      358 EAVNAFKELVRKYPQSPRARYGKAQCEDDDLAEKRSNEVLRGAIETTYQEVASLPDVPAD 417
Db      258 EAVNAFEELVRKYPQSPRARYGKAQCEDDDLAEKQSRNEVLRAIETTYQEAADLPDAPTD 317
QY      418 LKLSLKRSDRQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDNDAKKVY 477
Db      318 LVKLSLKRSEKQOFLGHRGSLTLQRLVQLFPNDTSLKNDLGVYLLIGDNDAKKVY 377
QY      478 BEVLSTPNDGFAKVHYGFIKAQNKIAESIPYLKEGIESGDPGTDGRFYFHLGDAMQV 537
Db      378 BEVLNTPNDGFAKVHYGFIKAQNKISESIPYLKEGIESGDPGTDGRFYFHLGDAMQV 437
QY      538 VGNKEAYKMYELGHRGFASVWQSLYNVNGLKAQPMWTPRETGYTELVKSLERNWKLI 597
Db      438 VGNKEAYKMYELGHRGFASVWQSLYNVNGLKAQPMWTPRETGYTELVKSLERNWKLI 497
QY      598 RDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKTCITLLEKPEET 657
Db      498 RDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKTCALLEKPEET 557
QY      658 TGCRRGQIKYSIMHPGTHWPHGTPTNCRRLMHLGLVIPKEGCKIRCANETRTWEEGKYL 717
Db      558 TGCRRGQIKYSIMHPGTHWPHGTPTNCRRLMHLGLVIPKEGCKIRCANETRTWEEGKYL 617
QY      718 IFDDSFHEVWQDASSFRLIFIVDVWHPBELTPQRRSLPAI 758
Db      618 IFDDSFHEVWQDASSFRLIFIVDVWHPBELTPQRRSLPAI 658

```

RESULT 6

```

Q9EQ66 PRELIMINARY; PRT; 689 AA.
ID Q9EQ66;
AC Q9EQ66;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
RT isoform of asph missing the catalytic domain share exons with
RT junction.";
RL J. Biol. Chem. 275:39543-39554 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Henderson N.L., Dinchuk J.E., Burn T.C., Hollis G.F., Friedman P.A.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289215; AAG39913.1; -.
DR EMBL; AF289205; AAG39913.1; JOINED.
DR EMBL; AF289206; AAG39913.1; JOINED.
DR EMBL; AF289207; AAG39913.1; JOINED.
DR EMBL; AF289208; AAG39913.1; JOINED.

```

DR EMBL; AF289209; AAG39913.1; JOINED.
DR EMBL; AF289210; AAG39913.1; JOINED.
DR EMBL; AF289211; AAG39913.1; JOINED.
DR EMBL; AF289212; AAG39913.1; JOINED.
DR EMBL; AF289213; AAG39913.1; JOINED.
DR EMBL; AF289214; AAG39913.1; JOINED.
DR MGD; MGI:1914186; Aspn.
DR InterPro; IPR007943; Asp-B-hydro N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 689 AA; 77319 MW; 9CB916DF109F432C CRC64;

Query Match 72.6%; Score 2919.5; DB 11; Length 689;
Best Local Similarity 74.7%; Pred. No. 5.2e-166;
Matches 573; Conservative 39; Mismatches 68; Indels 87; Gaps 9;

QY 1 MAQRKNAK-SSGNSSSSSGSGSGS-----TSAGSSSPGARRETKHGKNGRKGSLG 51
DB 1 MAPRKNAGGGGSSSSSGSGSGSGSPSTGSSSSSPGAR----- 43
QY 52 TSFTFWMTALLGWTSVAVWFDLVDYEEVLGKLGTYDADGDGDFVDVDAKVLGLKE 111
DB 44 -----GKLGTYDADGDGDFVDVDAKVLGLKE 70
QY 112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEFQNIIDEAKEQIOSLLHEMVHAHEVEGDLQO 171
DB 71 RSPSERTFPP-EAETHAELEEQAPEGADIQNEDEVKEQIOSLLQESVHTDH----DL-E 124
QY 172 EDGPTGEPOQDEDEFLMATDVDDRFETLBEVSHETEHSYHVEETVSQDCNQDMEMMS 231
DB 125 ADGLAGEPQPEVEDFLVTVDSDDRFEDLEPGTVHEIEEDTYHVEDTASQNHPNMDEMNTN 184
QY 232 EQENPDSSPEVEDERLHHTDDVTYQYVEEQAVYEPLENEGIEITEVTAPPEDNPEVDS 291
DB 185 EQNSDPSEAVTDAGVLLPHAEVVRHQDYDE-PYEPSEHEGVEIS-----DNITDDS 236
QY 292 QVIVEEVSIFPYEQEQEVPPEPTNRKTDDEQAKAVKKKKPKLTKFDKTIKAEIDAAEKL 351
DB 237 SIISEEINVASVEEQDTPP-----VKKKKPKLTKFDKTIKAEIDAAEKL 282
QY 352 RKRGIKEAVNAFKELVRKYPOSPPARYGKAQCEDDLAEKRSNEVLRGAIETVQEVASL 411
DB 283 RKRGIKEAVNAFEELVRKYPOSPPARYGKAQCEDDLAEKRSNEVLRAIETVQEVASL 342
QY 412 PDVPADLLKLSLKRSDRQQLGKNGSLTLQRLVQLFPENDTSLKNDLGVGYLIGDND 471
DB 343 PDAPTDLVKLSLKRSEERQQLGKNGSLTLQRLVQLFPESDTTLKNDLGVGYLIGDND 402
QY 472 NAKKVEEVLVTLPNDGFAKVHYGLLKAQNKAIESIPYKEGIESGDPGTDDGRFYFHL 531
DB 403 SAKKVEEVLNVTLPNDGFAKVHYGLLKAQNKAIESIPYKEGIESGDPGTDDGRFYFHL 462
QY 532 GDAMQVGNKEAYKMYELGKRGHPASVWQSLYVNGLKAQPMWTPKETGYTELVSLE 591
DB 463 GDAMQVGNKEAYKMYELGKRGHPASVWQSLYVNGLKAQPMWTPRETGYTELVSLE 522
QY 592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTTLL 651
DB 523 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLMQGRNENACKGAPKCTALL 582
QY 652 EKPETTGCRGQIKYSIMHPGTHVPHPTGPTNCRLRMHGLVTPKEGCKIRCANETRTW 711
DB 583 EKSETTGCRGQIKYSIMHPGTHVPHPTGPTNCRLRMHGLVTPKEGCKIRCANETRTW 642
QY 712 BEGKVLIFDSSFHEHFWODASSFRLLIFIVDVWHPDLTPQQRSLPAI 758
DB 643 BEGKVLIFDSSFHEHFWODASSFRLLIFIVDVWHPDLTPQQRSLPAI 689

Q9H2C4
ID Q9H2C4 PRELIMINARY; PRT; 313 AA.
AC Q9H2C4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Asparyl beta-hydroxylase 2.8 kb transcript.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
isoform of asph missing the catalytic domain share exons with
RT junction."
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF289489; AAG40811.1; -.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 313 AA; 34646 MW; 7885A18B1CD6D0D CRC64;

Query Match 40.7%; Score 1636; DB 4; Length 313;
Best Local Similarity 99.7%; Pred. No. 7.3e-90;
Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGNSSSSSGSGSGSTAGSSSPGARRETKHGKNGRKGSLGTSFTWFMV 60
DB 1 MAQRKNAKSSGNSSSSSGSGSGSTAGSSSPGARRETKHGKNGRKGSLGTSFTWFMV 60
QY 61 IALLGWTSVAVWFDLVDYEEVLGKLGTYDADGDGDFVDVDAKVLGLKERSTSEPAVP 120
DB 61 IALLGWTSVAVWFDLVDYEEVLGKLGTYDADGDGDFVDVDAKVLGLKERSTSEPAVP 120
QY 121 PEEAEPHTEPEEQVPVEAEFQNIIDEAKEQIOSLLHEMVHAHEVEGDLQOEDGPTGEPQ 180
DB 121 PEEAEPHTEPEEQVPVEAEFQNIIDEAKEQIOSLLHEMVHAHEVEGDLQOEDGPTGEPQ 180
QY 181 QEDDEFMATDVDDRFETLBEVSHETEHSYHVEETVSQDCNQDMEMMSQENPDSSSE 240
DB 181 QEDDEFMATDVDDRFETLBEVSHETEHSYHVEETVSQDCNQDMEMMSQENPDSSSE 240
QY 241 PVVEDERLHHTDDVTYQYVEEQAVYEPLENEGIEITEVTAPPEDNPEVDSQYIVEEVS 300
DB 241 PVVEDERLHHTDDVTYQYVEEQAVYEPLENEGIEITEVTAPPEDNPEVDSQYIVEEVS 300
QY 301 FPVEEQEQEVPPEP 313
DB 301 FPVEEQEQEVPPEP 313

RESULT 7

RESULT 8
Q9H291
ID Q9H291 PRELIMINARY; PRT; 299 AA.
AC Q9H291;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Junction.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564329; PubMed=11007777;
RA Treves S., Feriotto G., Moccagatta L., Gambari R., Zorzato F.;
RT "Molecular Cloning, Expression, Functional Characterization,

RT Chromosomal Localization, and Gene Structure of Junction, a Novel
RT Integral Calcium Binding Protein of Sarco(endo)plasmic Reticulum
RT Membrane."
RL J. Biol. Chem. 275:39555-39568(2000).
DR EMBL; AF306765; AAG42257.1; -.
DR GO; GO:0005789; C:cytoplasmic reticulum membrane; NAS.
DR GO; GO:0005509; F:calcium ion binding; NAS.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 299 AA; 33815 MW; 658F88C34EC2CA37 CRC64;

Query Match 36.2%; Score 1457.5; DB 4; Length 299;
Best Local Similarity 94.2%; Pred. No. 3.1e-79;
Matches 278; Conservative 2; Mismatches 0; Indels 15; Gaps 1;

QY 34 RETKHGKXGRKGGTSGTSPFTWMTALLGVWTSVAVWFVLDVYEEVL----- 84
Db 5 KETKHGKXGRKGGTSGTSPFTWMTALLGVWTSVAVWFVLDVYEEVLAKAKDPRYN 64
QY 85 -----GKLGIVDADGDGDFDVEDDAKVLGLKERSTSEPAVPPPEAEPHTEPEEQVVEA 138
Db 65 LSEVLQGLGIYDADGDGDFDVEDDAKVLGLKERSTSEPAVPPPEAEPHTEPEEQVVEA 124
QY 139 EPQNIIDEAKEQIQSLIHEMVHAHEVGEGLIQEDGPTGEPQEDDEFLMATDVDRPET 198
Db 125 EPQNIIDEAKEQIQSLIHEMVHAHEVGEGLIQEDGPTGEPQEDDEFLMATDVDRPET 184
QY 199 LEPEVSHETESHVHEETVSQDCNQDMEEMSEQENPDSSSEPVYEDERLHHTDVTYQ 258
Db 185 LEPEVSHETESHVHEETVSQDCNQDMEEMSEQENPDSSSEPVYEDERLHHTDVTYQ 244
QY 259 VYEEQAVYEPLNEGIEITEVTAPPEPNVEDSQVIVEVSTFPVEEQEVPPEP 313
Db 245 VYEEQAVYEPLNEGIEITEVTAPPEPNVEDSQVIVEVSTFPVEEQEVPPEP 299

RESULT 9

Q8TB28 PRELIMINARY; PRT; 270 AA.
AC Q8TB28;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Similar to aspartate beta-hydroxylase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025236; AAH25236.1; -.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 270 AA; 29757 MW; 8551773C7272202A CRC64;

Query Match 34.2%; Score 1376.5; DB 4; Length 270;
Best Local Similarity 85.9%; Pred. No. 1.8e-74;
Matches 269; Conservative 1; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAQRKNAKSSGSSSGSGSGTSGARRETKHGKXGRKGGTSGTSPFTWMT 60
Db 1 MAQRKNAKSSGSSSGSGSGTSGARRETKHGKXGRKGGTSGTSPFTWMT 60
QY 61 IALLGVWTSVAVWFVLDVYEEVLGKLGIVDADGDGDFDVEDDAKVLGLKERSTSEPAV 120
Db 61 IALLGVWTSVAVWFVLDVYEEVLGKLGIVDADGDGDFDVEDDAKVLGLKERSTSEPAV 120
QY 121 PEEAEPHTEPEEQVVEAEPQNIIDEAKEQIQSLIHEMVHAHEVGEGLIQEDGPTGEPQ 180
Db 121 PEEAEPHTEPEEQVVEAEPQNIIDEAKEQIQSLIHEMVHAHEVGEGLIQEDGPTGEPQ 163

QY 181 QEDDEFLMATDVDRPETLEPEVSHETESHVHEETVSQDCNQDMEEMSEQENPDSS 240
Db 164 -----ETESHVHEETVSQDCNQDMEEMSEQENPDSS 197
QY 241 PVVEDERLHHTDVTYQVVEEQAVYEPLNEGIEITEVTAPPEPNVEDSQVIVEVSI 300
Db 198 PVVEDERLHHTDVTYQVVEEQAVYEPLNEGIEITEVTAPPEPNVEDSQVIVEVSI 257
QY 301 FPVEEQEVPPEP 313
Db 258 FPVEEQEVPPEP 270

RESULT 10

Q9GQ82 PRELIMINARY; PRT; 785 AA.
AC Q9GQ82;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).
GN ASPH OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=20564328; Pubmed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction."
RL J. Biol. Chem. 275:39543-39554(2000).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Fianknoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan J.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

[illegible][illegible]

SQ SEQUENCE 308 AA; 33142 MW; F208C2C31C595282 CRC64;
Query Match
Best Local Similarity 24.8%; Score 998.5; DB 11; Length 308;
Matches 207; Conservative 30; Mismatches 62; Indels 23; Gaps 7;
QY 1 MAQRKNAK-SSGNSSSSSGSGS-----TSAGSSSPGARRETGHGKNGRKGGLSG 51
DB 1 MAPRKNAKGGGNGSSSSSGSGSGSPSTGSSSSSSPGARREAKHGKNGRRGGISG 60
QY 52 TSFPTWFMVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDADGDGDFDVKVLLGLKE 111
DB 61 GSFFTWMFVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDADGDGDFDVKVLLGLKE 120
QY 112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIIDEAKEQIQSLHVMVHAHVEGEDLQ 171
DB 121 RSPSEKTFPP-EAETHAELEEQAPGADIQNVDEVKEQIQSLQESVHTDH---DL-E 174
QY 172 EDGPTGEPOQEDDEFIMATDVDRFETLEPEVSHETESHVHEETVVSQDCNQDMEMMS 231
DB 175 ADGLAGEPOPEVEDEFLLVTDSDDRFEDLEPGTVHEIEDTYHVEDTASQNHPNDEMTN 234
QY 232 EQENPDSESPVEDERLHHTDDVTYQVVEEQAVYEPLNEGIEITETVTAAPPEDNVEDS 291
DB 235 EQENSDDPEAVTDAGVLLPHAEVVRHODYE-PVYEPSEHEGVEIS-----DNTIDDS 286
QY 292 QVIVEVSIFFVVEEQEVPPEP 313
DB 287 SIISEINVASVEEQDTPPDT 308
RESULT 12
Q91WG6 PRELIMINARY; PRT; 292 AA.
AC Q91WG6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Similar to aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015281; AAH15281.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 292 AA; 31568 MW; B3A470A46EC24B6D CRC64;
Query Match
Best Local Similarity 24.2%; Score 972.5; DB 11; Length 292;
Matches 202; Conservative 30; Mismatches 51; Indels 39; Gaps 8;
QY 1 MAQRKNAK-SSGNSSSSSGSGS-----TSAGSSSPGARRETGHGKNGRKGGLSG 51
DB 1 MAPRKNAKGGGNGSSSSSGSGSGSPSTGSSSSSSPGARREAKHGKNGRRGGISG 60
QY 52 TSFPTWFMVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDADGDGDFDVKVLLGLKE 111
DB 61 GSFFTWMFVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDADGDGDFDVKVLLGLKE 120
QY 112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIIDEAKEQIQSLHVMVHAHVEGEDLQ 171
DB 121 RSPSEKTFPP-EAETHAELEEQAPGADIQNVDEVKEQIQSLQESVHTDH---DL-E 174
QY 172 EDGPTGEPOQEDDEFIMATDVDRFETLEPEVSHETESHVHEETVVSQDCNQDMEMMS 231
DB 175 ADGLAGEPOPEVEDEFLLVTDSDDRFEDLEPGTVHEIEDTYHVEDTASQNHPNDEMTN 234
QY 232 EQENPDSESPVEDERLHHTDDVTYQVVEEQAVYEPLNEGIEITETVTAAPPEDNVEDS 291
DB 235 EQENSDDPEAVTDAGVLLPHAEVVRHODYE-PVYEPSEHEGVEIS-----DNTIDDS 286
QY 292 QVIVEVSIFFVVEEQEVPPEP 313
DB 287 SIISEINVASVEEQDTPPDT 308

DB 175 ADGLAGEPOPEVEDEFLLVTDSDDRFEDLEPGTVHEIEDTYHVEDTASQNHPNDEMTN 234
QY 232 EQENPDSESPVEDERLHHTDDVTYQVVEEQAVYEPLNEGIEITETVTAAPPEDNVEDS 291
DB 235 EQEN-----SEEVVRHODYE-PVYEPSEHEGVEIS-----DNTIDDS 270
QY 292 QVIVEVSIFFVVEEQEVPPEP 313
DB 271 SIISEINVASVEEQDTPPDT 292
RESULT 13
Q920F9 PRELIMINARY; PRT; 270 AA.
ID Q920F9;
AC Q920F9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Cardiac junctate 1.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Hong C.-S., Kim D.H.;
RT "Cloning of mouse junction homologs.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF302653; AAL09319.2; -.
DR PIR; JC7792; JC7792.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 270 AA; 29949 MW; 70B13DC01B51701 CRC64;
Query Match
Best Local Similarity 22.1%; Score 890; DB 11; Length 270;
Matches 179; Conservative 30; Mismatches 57; Indels 14; Gaps 5;
QY 34 RETKHGKNGRKGGLSGTSFPTWFMVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDAD 93
DB 5 KEAKHGKNGRRGGISGSGSFPTWFMVIALGVWTSVAVVMFDLVDYEEVLGKLGIVDAD 64
QY 94 GDGDFVDVDAKVLGLKERSTSEPAVPPEEAEPHTEPEEQVVEAEPQNIIDEAKEQIQS 153
DB 65 GDGDFVDVDAKVLGLKERSPSERTFPP-EAETHAELEEQAPGADIQNVDEVKEQIQS 123
QY 154 LHHMVAHVEGEDLQQEDGPTGEPOQEDDEFIMATDVDRFETLEPEVSHETESHYH 213
DB 124 LQESVHTDH---DL-EADGLAGEPOPEVEDEFLLVTDSDDRFEDLEPGTVHEIEDTYH 178
QY 214 VEETVSQDCNQDMEMMSSEQENPDSESPVEDERLHHTDDVTYQVVEEQAVYEPLNEG 273
DB 179 VEDTASQNHPNDEMTNEQENSDDPEAVTDAGVLLPHAEVVRHODYE-PVYEPSEHSG 237
QY 274 IEITVTAAPPEDNVEDSQVIVEVSIFFVVEEQEVPPEP 313
DB 238 VEIS-----DNTIDDSIISEINVASVEEQDTPPDT 270
RESULT 14
Q93178 PRELIMINARY; PRT; 872 AA.
ID Q93178;
AC Q93178; Q93876;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE K09A9.6 protein.
GN K09A9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhaditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swindburne J.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z79601; CAB01887.1; -.
DR EMBL; Z79596; CAB01887.1; JOINED.
DR EMBL; Z79596; CAB01859.1; -.
DR EMBL; Z79601; CAB01859.1; JOINED.
DR PIR; T18861; T18861.
DR WormRep; K09A9.6; CE11982.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 872 AA; 99456 MW; 8CA3E03489A9848B CRC64;

Query Match 19.8%; Score 798; DB 5; Length 872;
Best Local Similarity 26.8%; Pred. No. 3.1e-39;
Matches 191; Conservative 144; Mismatches 278; Indels 100; Gaps 15;

QY 89 TYDADGGDFVDVDAKVLGLKERSTSEPAVPPEBAEPHTEPEEQ-----VPEAEPO 141
DB 205 VEDDDDDDDDDDDVE-----APAPQEPSRKQKANHQRREEKDKNKKYQPYKEEPD 255

QY 142 NIEDEAKEQIQSLHEMVHAHEGEGDLQOEDG-----PTGEFQOEDDEFMATDVDR 195
DB 256 LDDDDDDDDDDDDGDDDDDEKDDGDAEKDDGDDDDDDDDGNDDEEDGKNKKTVEAK 315

QY 196 FETL-----EPEVSHETE--HSYHVEET-----VSQDC-NQDMEE- 228
DB 316 SDKVVEEKNDGDEPHVSHREARQLRQHLRESRDNRPRQGNRECIHQDCPNRESLKP 375

QY 229 ----NMSEQENPDSEFPVEDERLHHTDDVTYQVVEEQAVYEPLNEGIEITEVTAPPE 284
DB 376 RKSLLVTKKTKSMVERILDDDEFDEDDDD-----EDDSEVNK-----N 415

QY 285 DNPVEDSQVIVEEVSIFPVEEQQVPEPTNKTDDPEQAKAVKKKPKLTKEDTKIAE 344
DB 416 DREDDDEDVDERIS-----DRUSSSSSYKRHAITTKETIGFRDI 454

QY 345 LDAAEKLRRKRIEAVNAFKELVRKYPQSPRARYGKACEDDLAEKRSNEVLGALET 404
DB 455 LDRADNLVEKHQYEEAMELFDHVIAVYPASTRAYFGKARAYDIRGEIAEDETDRDKAIEI 514

QY 405 YQEVASLPVPAQLLKLKRSRSDROFLGMRGSLTLQRLVQLFPNDTSLKNDLGVS 464
DB 515 YEKILQNSGVDPALFRQAQRLIEKTRFRGLHKTTLTAHYFIDRPEELNLQDPAISF 574

QY 465 LLIGDNDNAKYYEVLVTPNDGPAKVHYGFIKA-QNKIAESIPYLKEGIESGDPGTD 523
DB 575 VMKRYEDARTVLKNVLANDPNHVALAYGYILKAHDKVEQGVALLMRKSLKNADNEIT 634

QY 524 DGRFYFHLGDAMQVRGNK-EAYKMYELGKRGHFAVWQRLYNVNGIKAQPPWTPKETG 582
DB 635 DPKFYIQLGHGLTLTGKSEADAVYQKAAQMGVFMFTAQQRSLYNIEGLTGRAWMAMDQTP 694

QY 583 YTELVSLEARNMKLIRDEGLAVMDKAGLFLPEDENLREKGDWSQFTLMQGRNENACK 642
DB 695 YSKFLKTVERQWATIRQEGMEVLKDCSDCWLHDNQQLVIDGQWKFPIMSEQNFIKSSCE 754

QY 643 GAPKCTLLEKFPETTGCRGQIKYSIMHPGTHVPHPTGPTNCRRLRMHLGLVIPKEGCKI 702
DB 755 RMPQCTLIQEFPAASSNASKSDMLSVLSSGASILPHCGPTNYHLQAHGLVSPSE-ARI 813

QY 703 RCANETRTWEEKVLIFDDSEHEVWOD---ASSFLFIVDVWHPELTPQOR 752
DB 814 RVGNETKWRSGKFTIYDSEFHEHLQFDGASSSSFRVLTLTQLMHPVQPHQR 866

RESULT 15
Q920F8 PRELIMINARY; PRT; 259 AA.
AC Q920F8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cardiac junctate 2.
GN ASPH OR 3110001L23RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Hong C.-S., Kim D.H.;
RT "Cloning of mouse junction homologs.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF302654; AAL09320.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 259 AA; 28455 MW; A6740C6CF199E093 CRC64;

Query Match 19.4%; Score 779.5; DB 11; Length 259;
Best Local Similarity 55.4%; Pred. No. 7.5e-39;
Matches 163; Conservative 27; Mismatches 51; Indels 53; Gaps 6;

QY 34 RETKHGHKNGRKGLSGSTFTWFMVIALIGVWTSVAVVWEDLVDEEVLGKLGTYDAD 93
DB 5 KEAKHGHHKNGRRGISGGSFTWFMVIALIGVWTSVAVVWFDLVDEEVLGKLGTYDAD 64

QY 94 GDGDFVDDAKVLL-----GLKERSTSEPAVPPEBAEPHTEPEEQVPEAE 139
DB 65 GDGDFVDDAKVLLLEGGGLAKRKTAKGLKERSPSERTPP-EAETHAELEEQAD---- 119

QY 140 PQNIDEAKEQIQSLHEMVHAHEGEGDLQOEDGPTGEFQOEDDEFMATDVDRFETL 199
DB 120 -----EGADL-EADGLAGEPQPEVEDFLVTITDSDRFEDL 153

QY 200 EPEVSHETEHSYHVEETVSQDCNQDMEEMSEQENPDSEFPVEDERLHHTDDVTYQV 259
DB 154 EPGTVHEIEDTYHVEDTASQNHPNMDEEMTNEQENSDPSEAVTDAGVLLPHAEVVRHQD 213

QY 260 YEEQAVYEPLNEGIEITEVTAPBEDNPVEDSQVIVEEVSIFPVEEQQVPEPT 313
DB 214 YDE-PYVERSEHEGVEIS-----DNTIIDSIISEIINVASVEEQDTPPDPT 259

Search completed: May 19, 2004, 15:59:50
Job time : 76.0139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:50:28 ; Search time 4.0076 Seconds
(without alignments)
2185.588 Million cell updates/sec

Title: US-09-903-199-4
Perfect score: 139
Sequence: 1 CDXXXXXXKXGNGXCDXXCNNAACXXDXGDC 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	115	82.7	31	4	AAB83920	Aab83920 Amino aci
2	115	82.7	31	6	ABG72366	Abg72366 Epidermal
3	115	82.7	31	6	ADA00641	Ada00641 EGF-like
4	105	75.5	2703	4	ABB60266	Abb60266 Drosophil
5	85	61.2	2531	7	ADE63713	Aag79774 Rat Prote
6	85	61.2	2531	7	ADE63705	Ade63705 Rat Prote
7	85	61.2	2531	7	ADE63709	Ade63709 Rat Prote
8	85	61.2	2531	7	ADE63701	Ade63701 Rat Prote
9	84	60.4	2444	5	ABB07821	Abb07821 Constitut
10	84	60.4	2469	5	AAE18207	Aae18207 Human MOL
11	84	60.4	2469	7	ADD18192	Add18192 Human mol
12	84	60.4	2471	2	AAO27065	Aao27065 Human Not
13	84	60.4	2471	2	AAV06816	Aay06816 Human Not
14	84	60.4	2471	6	AAG79774	Aag79774 Human Not
15	84	60.4	2471	6	ABP72572	Abp72572 Human Not
16	84	60.4	2471	6	ABR61831	Abtr61831 Human Not
17	84	60.4	2471	7	ABR61760	Abtr61760 Human Not
18	84	60.4	2471	7	ADE58245	Ade58245 Human Pro
19	84	60.4	2471	7	ADE63707	Ade63707 Human Pro
20	84	60.4	2471	7	ADE63703	Ade63703 Human Pro
21	84	60.4	2471	7	ADE63715	Ade63715 Human Pro
22	84	60.4	2471	7	ADE58243	Ade58243 Rat Prote
23	84	60.4	2471	7	ADE63711	Ade63711 Human Pro
24	84	60.4	2556	2	AAO27066	Aao27066 Human Not
25	84	60.4	2556	6	ABG70518	Abg70518 Human pol

26	84	60.4	2556	6	AAG79773	Aag79773 Human Not
27	84	60.4	2556	6	ABP72571	Abp72571 Human Not
28	84	60.4	2556	6	ABR61830	Abtr61830 Human Not
29	84	60.4	2556	7	ABR61759	Abtr61759 Human Not
30	83	59.7	1872	2	AAW68510	Aaw68510 Partial h
31	83	59.7	2321	2	AAW49698	Aaw49698 Human Not
32	79	56.8	1964	2	AAW95557	Aaw95557 Mus muscu
33	78	56.1	178	6	ABU11844	Abu11844 Human MDD
34	78	56.1	846	6	ADA54824	Ada54824 Human pro
35	78	56.1	928	4	ABR61376	Abtr61376 Human Glc
36	78	56.1	928	7	ADD27812	Add27812 Human Glc
37	78	56.1	928	7	ABW01488	Abw01488 Human Glc
38	78	56.1	928	7	ABW01537	Abw01537 Human Glc
39	78	56.1	1199	7	ADD27810	Add27810 Soluble h
40	78	56.1	1199	7	ABW01487	Abw01487 N-acetyl g
41	78	56.1	1199	7	ABW01536	Abw01536 N-acetyl g
42	78	56.1	1256	5	AAE25290	Aae25290 Human nuc
43	78	56.1	1459	6	ABU07381	Abu07381 Human pro
44	77	55.4	1078	2	AAE28963	Aae28963 Notch hn3
45	77	55.4	1078	2	AAO27058	Aao27058 Human Not

ALIGNMENTS

RESULT 1	
AAB83920	
ID AAB83920 standard; peptide; 31 AA.	
XX	
AC AAB83920;	
XX	
DT 23-JUL-2001 (first entry)	
XX	
DE Amino acid sequence of consensus epidermal growth factor-like domain.	
XX	
KW Epidermal growth factor-like domain; EGF-like domain; cancer;	
KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.	
XX	
OS Synthetic.	
XX	
FH Key	Location/Qualifiers
FT Misc-difference 1..31	/note= "Xaa represents an unspecified residue"
XX	
PN WO200135102-A2.	
XX	
PD 17-MAY-2001.	
XX	
PF 08-NOV-2000; 2000WO-US030738.	
XX	
PR 08-NOV-1999; 99US-00436184.	
XX	
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.	
XX	
PI Wands JR, De La Monte SM, Ince N, Carlson RI;	
XX	
DR WPI; 2001-329171/34.	
XX	
PT Diagnosing malignant neoplasm in a mammal, involves contacting mammalian	
PT sample with antibody that binds to human aspartyl beta-hydroxylase	
PT polypeptide to form antigen-antibody complex and detecting the complex.	
XX	
PS Disclosure; Page 7; 76pp; English.	
XX	
CC The present sequence represents a consensus epidermal growth factor (EGF)	
CC -like cysteine-rich repeat. EGF-like domains of polypeptides are	
CC hydroxylated by a human aspartyl (asparaginyl) beta-hydroxylase (HAAH)	
CC enzyme. HAAH is used in the method of the invention. The specification	
CC describes a method for diagnosing a malignant neoplasm in a mammal. The	
CC method comprises contacting a body fluid with an antibody which binds to	
CC HAAH polypeptide under complex forming conditions, and detecting the	
CC antigen-antibody complex. The method is useful for diagnosing and	
CC prognosing a malignant neoplasm in a bodily fluid e.g. central nervous	

CC system (CNS)-derived body fluid, blood, serum, urine, saliva, sputum,
CC lung effusion, and ascites fluid of mammal, where the neoplasm is derived
CC from endodermal tissue and is selected from colon cancer, breast cancer,
CC pancreatic cancer, liver cancer, cancer of bile ducts, and cancer of the
CC CNS. HAAH antibodies, linked to a cytotoxic agent, are useful for killing
CC tumour cells
XX
SQ Sequence 31 AA;

Query Match 82.7%; Score 115; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31

RESULT 2
ABG72366 ID ABG72366 standard; peptide; 31 AA.
XX AC ABG72366;
XX DT 06-FEB-2003 (first entry)
XX DE Epidermal growth factor-like cyteine-rich repeat consensus sequence.
XX
KM Aspartyl (asparaginyl) beta-hydroxylase; HAAH; cytosstatic;
KM immunostimulant; antibody; neoplasm; tumour; FB50; 86A; 5C7; 19B;
KM brain tumour; glioma; glioblastoma; astrocytoma; haemangioma;
KM pancreatic carcinoma; colon cancer; breast cancer; pancreatic cancer;
KM liver cancer; cancer of the bile ducts; primary malignant CNS neoplasm;
KM metastatic CNS neoplasm; EGF; NOTCH;
KM Epidermal growth factor-like cyteine-rich repeat consensus sequence.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 3..5
FT Misc-difference /label= UNKNOWN
FT Misc-difference 7..8
FT Misc-difference /label= UNKNOWN
FT Misc-difference 10
FT Misc-difference /label= UNKNOWN
FT Misc-difference 14
FT Misc-difference /label= UNKNOWN
FT Misc-difference 17..18
FT Misc-difference /label= UNKNOWN
FT Misc-difference 25..26
FT Misc-difference /label= UNKNOWN
FT Misc-difference 29
FT Misc-difference /label= UNKNOWN
XX
PN US2002110559-A1.
XX
PD 15-AUG-2002.
XX
PF 17-MAY-2001; 2001US-00859604.
XX
PR 08-NOV-1999; 99US-00436184.
XX
PA (WAND/) WANDS J R.
PA (DMON/) DE LA MONTE S M.
PA (DEUT/) DEUTCH A H.
PA (GHAN/) GHANBARI H A.
XX
PI Wands JR, De la Monte SM, Deutch AH, Ghanbari HA;
XX
DR WPI; 2003-066676/06.
XX
FT Diagnosing neoplasm and inhibiting tumor growth, by contacting tissue of
FT mammal with detectably-labeled antibody which binds to human aspartyl

PT (asparaginyl) beta-hydroxylase.
XX
PS Disclosure; Page 5; 34pp; English.
XX
CC The invention relates to diagnosing a neoplasm and inhibiting tumour
CC growth in a mammal, using an antibody that binds to human aspartyl
CC (asparaginyl) beta-hydroxylase (HAAH). Diagnosing a neoplasm comprises
CC contacting a tissue with a detectably-labelled antibody where an increase
CC in level of antibody binding at tissue site compared to the level of
CC binding to normal non-neoplastic tissue indicates the presence of a
CC neoplasm at the tissue site. Inhibiting tumour growth in mammal involves
CC administering the antibody conjugated to a cytotoxic agent to a mammal.
CC Also included are a method of conferring an immune response to a tumour
CC cell in a mammal, by administering the antibody, a method of inducing an
CC HAAH-specific immune response in a mammal, by administering to the mammal
CC an HAAH polypeptide (or a polynucleotide composition encoding the
CC polypeptide, or its degenerate variant), a fragment of HAAH comprising an
CC extracellular domain and lacking a cytoplasmic domain of HAAH, an
CC antibody or its fragment which binds to HAAH (where the antibody is FB50,
CC 86A, 5C7 or 19B), a hybridoma cell line chosen from hybridoma FB50,
CC HA386A, HA15CA and HA219B, and a fragment of HAAH which lacks enzymatic
CC activity or alpha-ketoglutarate binding domain and epidermal growth
CC factor (EGF)-like domain. The methods are useful for diagnosing neoplasm
CC in a mammal, inhibiting tumour growth in a mammal, conferring an immune
CC response to a brain tumour cell (e.g. glioma, glioblastoma, astrocytoma
CC or haemangioma) in a mammal, for conferring immune response to a
CC pancreatic carcinoma cell and for inducing a HAAH-specific immune
CC response in a mammal. The method is useful for diagnosing malignant
CC neoplasms derived from endodermal tissue, e.g. colon cancer, breast
CC cancer, pancreatic cancer, liver cancer and cancer of the bile ducts. The
CC method is also useful for diagnosing neoplasms of central nervous system
CC (CNS) e.g. primary malignant CNS neoplasms of both neuronal and glial
CC cell origin and metastatic CNS neoplasms, and for diagnosing brain
CC tumours e.g. glioma, glioblastoma, astrocytoma or haemangioma. The
CC present sequence represents an Epidermal growth factor (EGF)-like cyteine
CC -rich repeat consensus sequence (e.g. from the \NOTCH protein) which is
CC hydroxylated by human HAAH
XX
SQ Sequence 31 AA;

Query Match 82.7%; Score 115; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31

RESULT 3
ADA00641 ID ADA00641 standard; peptide; 31 AA.
XX AC ADA00641;
XX
DT 06-NOV-2003 (first entry)
XX
DE EGF-like cysteine-rich repeat consensus sequence.
XX
KM Tumour growth inhibition; human aspartyl (asparaginyl) beta-hydroxylase;
KM HAAH hydroxylation; NOTCH polypeptide;
KM epidermal growth factor-like repeat; EGF-like repeat; tumour cell;
KM malignant neoplasm; colon cancer; breast cancer; pancreatic cancer;
KM liver cancer; cancer of the bile duct; cancer the central nervous system;
KM CNS; cytosstatic; EGF-like cysteine-rich repeat.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 3..5
FT Misc-difference /label= Unknown
FT FT Misc-difference 7..8
FT FT Misc-difference /label= Unknown

FT Misc-difference 10 /label= unknown
FT Misc-difference 14 /label= unknown
FT Misc-difference 17.18 /label= unknown
FT Misc-difference 25.26 /label= unknown
FT Misc-difference 29 /label= unknown
FT Misc-difference 29 /label= unknown
XX US2003031670-A1.
XX 13-FEB-2003.
XX 08-NOV-1999; 99US-00436184.
XX 08-NOV-1999; 99US-00436184.
XX 08-NOV-1999; 99US-00436184.
XX (WAND/) WANDS J R.
XX (DMON/) DE LA MONTE S M.
XX (INCE/) INCE N.
XX (CARL/) CARLSON R I.
XX Wands JR, De La Monte SM, Ince N, Carlson RI;
XX WPI; 2003-605701/57.
XX Inhibiting tumor growth or killing tumor cells (e.g. cancer of the colon,
PT breast, pancreatic, liver or the central nervous system), by
PT administering an inhibitor of the human aspartyl (asparaginyl) beta-
PT hydroxylase.
XX Disclosure; Page 4; 30pp; English.
XX The present invention relates to a method for inhibiting tumour growth in
CC a mammal. The method comprises administering to the mammal a compound,
CC which inhibits the expression or enzymatic activity of a human aspartyl
CC (asparaginyl) beta-hydroxylase (HAAH). The compound may inhibit HAAH
CC hydroxylation of a NOTCH polypeptide. In particular, the compound may
CC inhibit hydroxylation of an epidermal growth factor (EGF)-like repeat
CC sequence in a NOTCH polypeptide. The methods are useful for inhibiting
CC tumour growth or killing tumour cells, or for diagnosing or
CC prognosticating a malignant neoplasm. In particular, the tumour or
CC neoplasm is colon cancer, breast cancer, pancreatic cancer, liver cancer,
CC cancer of the bile ducts, or cancer or tumour of the central nervous
CC system (CNS). The present sequence represents an EGF-like cysteine-rich
CC repeat consensus sequence.
XX Sequence 31 AA;
SQ

Query Match 82.7%; Score 115; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 7.1e-08;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDXXXXXXKXGXGXCXXXCNNAACXXDXDC 31
Db 1 CDXXXXXXKXGXGXCXXXCNNAACXXDXDC 31

RESULT 4
ABB60266
ID ABB60266 standard; protein; 2703 AA.
XX ABB60266;
AC ABB60266;
XX 26-MAR-2002 (first entry)
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 7590.
DE Drosophila melanogaster polypeptide SEQ ID NO 7590.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX

OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EM;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL04369.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 7590; 21pp + Sequence listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2703 AA;
SQ

Query Match 75.5%; Score 105; DB 4; Length 2703;
Best Local Similarity 54.8%; Pred. No. 5.3e-05;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXXXXKXGXGXCXXXCNNAACXXDXDC 31
Db 1482 CDRGCTEKQGNIGCDSDCNTTYACNFDGNDG 1512

RESULT 5
ADE63713
ID ADE63713 standard; protein; 2531 AA.
XX ADE63713;
AC ADE63713;
XX 29-JAN-2004 (first entry)
DT 29-JAN-2004 (first entry)
XX Rat Protein CAA40667, SEQ ID NO 9657.
DE Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX Rattus norvegicus.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GEHO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
PA

XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX PA (GENO) GEN HOSPITAL CORP.
DR DR WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX PI New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2531 AA;
QY Query Match 61.2%; Score 85; DB 7; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
DB 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
1449 CELPECQEDAGNKVCNLTQCNNHACGWDGDC 1479
RESULT 6
ADE63705
ID ADE63705 standard; protein; 2531 AA.
XX AC ADE63705;
XX DT 29-JAN-2004 (first entry)
XX DE Rat Protein CAA40667, SEQ ID NO 9649.
XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.
XX (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX DR WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2531 AA;
QY Query Match 61.2%; Score 85; DB 7; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
DB 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
1449 CELPECQEDAGNKVCNLTQCNNHACGWDGDC 1479
RESULT 7
ADE63709
ID ADE63709 standard; protein; 2531 AA.
XX AC ADE63709;
XX DT 29-JAN-2004 (first entry)
XX DE Rat Protein CAA40667, SEQ ID NO 9653.
XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2531 AA;
Query Match 61.2%; Score 85; DB 7; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
QY 1 CDXXXCXXKXGNGXCDXXCNNNAACXXDXDC 31
DB 1449 CELPECQEDAGNKVCNLCNNHACGWDGDC 1479
RESULT 8
ADE63701
ID ADE63701 standard; protein; 2531 AA.
XX
AC ADE63701;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein CAA40667, SEQ ID NO 9645.
XX
KM Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KM chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX

PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2531 AA;
Query Match 61.2%; Score 85; DB 7; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
QY 1 CDXXXCXXKXGNGXCDXXCNNNAACXXDXDC 31
DB 1449 CELPECQEDAGNKVCNLCNNHACGWDGDC 1479
RESULT 9
ABB07821
ID ABB07821 standard; protein; 2444 AA.
XX
AC ABB07821;
XX
DT 03-JUL-2002 (first entry)
XX
DE Constitutively active notch-1 protein.
XX
KM Cell differentiation; notch; epidermis; cytosratic; dermatological;
XX

```

KW epithelial; skin; cancer; gamma secretase; notch-1.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 891 /label= Gly or Ala
FT /note= "encoded by GSS"
FT Misc-difference 1763 /label= Glu, Arg, Pro or Leu
FT /note= "encoded by CNG"
FT Misc-difference 1787 /label= Thr, Ala, Pro or Ser
FT /note= "encoded by NCC"
XX
PN WO200218544-A2.
XX
PD 07-MAR-2002.
XX
PF 31-AUG-2001; 2001WO-US027246.
XX
PR 31-AUG-2000; 2000US-0229614P.
XX
PA (LOYO ) UNIV LOYOLA CHICAGO.
XX
PI Nickolooff BJ, Miele L;
XX
DR WPI; 2002-339659/37.
DR N-PSDB; ABL40768.
XX
PT Inducing differentiation of epithelial cell useful for inducing barrier
PT formation within epithelium for treating psoriasis, sunburn, involves
PT exogenously providing a source of a Notch agonist to the epithelial cell.
XX
PS Claim 10; Page 50-66; 101pp; English.
XX
CC The invention relates to a method of inducing differentiation of at least
CC one epithelial cell. The method involves exogenously providing at least
CC one source of at least one Notch agonist to at least one epithelial cell,
CC whereby the Notch pathway is activated within at least one epithelial
CC cell so that the differentiated epidermis; for assaying for genetic propensity
CC of a patient to develop a disorder associated with epithelial barrier
CC formation; for retarding progression of skin cancer and for diagnosing
CC aggressive melanoma are also provided. The methods are useful for
CC inducing differentiation of at least one epithelial cell e.g. a
CC keratinocyte or a pre-malignant cell, in vivo or ex vivo. The method is
CC useful for inducing differentiation of epithelial cell within cutaneous
CC epithelial tissue or dermal equivalent, or within extracutaneous
CC tissue, gastrointestinal epithelia, urogenital epithelia, or respiratory
CC epithelia. The methods are useful retarding the progression of skin
CC cancer such as aggressive melanoma, aggressive cutaneous T-cell lymphoma
CC (CTCL), aggressive squamous cell carcinoma, or aggressive basal cell
CC carcinoma, by preferably administering an antagonist of the Notch pathway
CC such as gamma secretase inhibitor. The present sequence represents a
CC constitutively active notch-1 protein
XX
SQ Sequence 2444 AA;
Query Match 60.4%; Score 84; DB 5; Length 2444;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;
Db 1 CDXXXXXXXGXGXCDDXNNNAACXXDXDC 31
: : | | | | | | | | | |
: : | | | | | | | | | |
1450 CELPBCQEDAGNKVCSLQCNNHACGWDGGDC 1480
RESULT 10
AAE18207
ID AAE18207 standard; protein; 2469 AA.
XX

```

AC	AAE18207;
XX	
DT	07-MAY-2002 (first entry)
XX	
DE	Human MOL1a protein.
XX	
KW	Secreted molecule; MOL1a protein; MOLX; cardiomyopathy; atherosclerosis;
KW	diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;
KW	liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
KW	cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
KW	immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;
KW	HIV; human immunodeficiency virus; hepatitis; haemostatic disease; pain;
KW	haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;
KW	Parkinson's disease; Huntington's disease; muscular disease; stress;
KW	ocular disease; growth disorder; depression; epilepsy; contraceptive;
KW	vulnerary; osteopathic; haemostatic; tranquiliser; antidepressant;
KW	analgesic; vasotropic; hypotensive; gene therapy.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	Peptide
FT	1..25
FT	/label= Signal_peptide
FT	Protein
FT	26..2469
FT	/label= Mature_MOL1a_protein
XX	
PN	WO200206339-A2.
XX	
PD	24-JAN-2002.
XX	
PF	03-JUL-2001; 2001WO-US021249.
XX	
PR	03-JUL-2000; 2000US-0215854P.
PR	03-JUL-2000; 2000US-0215856P.
PR	03-JUL-2000; 2000US-0215902P.
PR	07-JUL-2000; 2000US-0216585P.
PR	07-JUL-2000; 2000US-0216586P.
PR	07-JUL-2000; 2000US-0216722P.
PR	17-JUL-2000; 2000US-0218622P.
PR	17-JUL-2000; 2000US-0218992P.
PR	27-JUL-2000; 2000US-0221285P.
PR	14-FEB-2001; 2001US-0268734P.
PR	08-MAR-2001; 2001US-0274260P.
PR	29-MAR-2001; 2001US-0279856P.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;
PI	Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;
PI	Alsbrook J, Lepley DM, Shen L, Burgess CE, Shimkets RA, Padigaru M;
XX	
DR	WPI: 2002-155038/20.
DR	N-PSDB; AAD28941.
XX	
PT	Nucleic acids encoding secreted polypeptides, designated MOLX
PT	polypeptides, useful for treating a MOLX-associated disorder, e.g.
PT	cardiomyopathy, atherosclerosis, diabetes and metabolic disorders.
XX	
PS	Claim 1; Page 11; 223pp; English.
XX	
CC	The patent discloses nucleic acid sequences encoding novel secreted
CC	molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL
CC	protein where X is an integer from 1 to 8). Sequences of the invention
CC	are useful for treating or preventing a MOLX-associated disorder in
CC	humans. They are useful for treating or preventing cardiomyopathy,
CC	atherosclerosis and disorders related to cell signal processing and
CC	metabolic pathway modulation. The MOLX antibodies are useful for treating
CC	or preventing diabetes and disorders related to cell signal processing
CC	and metabolic pathway modulation. MOLX sequences are useful for the
CC	treatment or diagnosis of other MOLX-associated disorders, e.g.
CC	chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
CC	scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
CC	thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune

CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,
CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,
CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC ocular disease, muscular diseases, growth disorders, loss of libido,
CC stress, depression, pain and epilepsy. They are useful for preventing
CC chemotherapy side effects and as contraceptives. Sequences of the
CC invention are also useful for gene therapy. The present sequence is human
CC Notch-like protein, MOL1a
XX
SQ Sequence 2469 AA;

Query Match 60.4%; Score 84; DB 5; Length 2469;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1423 CLSQYCADKARDGVCDACNSHACQWDGGDC 1453

RESULT 11

ADD18192
ID ADD18192 standard; protein; 2469 AA.

XX
AC ADD18192;

XX
DT 15-JAN-2004 (first entry)

XX
DE Human molecule (MOL) protein MOL1a.

XX
KW molecule protein; MOL protein; MOLX; MOLX agonist; MOLX antagonist;
KW cardiant; antidiabetic; antiarteriosclerotic; gene therapy;
KW MOLX-associated disorder; cardiomyopathy; diabetes; atherosclerosis;
KW human; MOL1a.

XX
OS Homo sapiens.

XX
PN WO2003003984-A2.

XX
PD 16-JAN-2003.

XX
PF 03-JUL-2002; 2002WO-US021268.

XX
PR 05-JUL-2001; 2001US-0303168P.

XX
PR 05-JUL-2001; 2001US-0303241P.

XX
PR 26-SEP-2001; 2001US-00965212.

XX
PR 26-SEP-2001; 2001US-00966545.

XX
PR 26-SEP-2001; 2001US-00966546.

XX
PR 01-APR-2002; 2002US-0368996P.

XX
PR 01-APR-2002; 2002US-0369065P.

XX
PR 08-MAY-2002; 2002US-0378730P.

XX
PR 30-MAY-2002; 2002US-0384327P.

XX
PR 07-JUN-2002; 2002US-0386816P.

XX
PR 17-JUN-2002; 2002US-00174372.

XX
PA (CURA-) CURAGEN CORP.

XX
PI Fernandes ER, Vernet CAM, Shimkets RA, Anderson DW, Padigaru M;
PI Boldog FL, Li L, Shenoy SG, Casman SJ, Rastelli L, Alsobrook JP;
PI Burgess CE, Grosse WM, Gusev VY, Ji W, Lepley DM, Liu X, Mezick AJ;
PI Patturajan M, Shen L, Spaderna SK, Spytek KA, Szekeres ES;
PI Taupier RJ, Tchernev VT, Zernusen BD, Voss EZ;

XX
DR WPI; 2003-210304/20.
DR N-PSDB; ADD18191.

XX
PT New MOLX polypeptide, nucleic acid or MOLX-specific antibody, useful for
PT preparing a composition for treating or preventing a MOLX-associated
PT disorder, e.g., cardiomyopathy, diabetes or atherosclerosis.

XX
PS Claim 1; SEQ ID NO 2; 371pp; English.

CC This invention relates to novel human nucleic acid sequences which encode
CC novel molecule (MOL) proteins numbered MOL1-23, referred to generally in
CC the specification as MOLX. Compounds which modulate the function of the
CC MOLX proteins of the invention, MOLX agonists or antagonists, may have
CC cardiant, antidiabetic or antiarteriosclerotic activities. In addition,
CC the DNA and protein sequences disclosed may prove useful for gene
CC therapy. The protein, nucleic acid or antibody is useful for preparing a
CC composition for treating or preventing a MOLX-associated disorder, for
CC example cardiomyopathy, diabetes or atherosclerosis. The present sequence
CC is the amino acid sequence of a MOL protein of the invention.
XX
SQ Sequence 2469 AA;

Query Match 60.4%; Score 84; DB 7; Length 2469;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1423 CLSQYCADKARDGVCDACNSHACQWDGGDC 1453

RESULT 12

AAO27065
ID AAO27065 standard; protein; 2471 AA.

XX
AC AAO27065;

XX
DT 22-MAY-2003 (first entry)

XX
DE Human Notch protein homologue sequence, SEQ ID No 19.

XX
KW Notch protein; cell fat disorder; differentiation; cancer; breast; colon;
KW cervical cancer; pre-neoplastic; non-malignant state; neoplastic;
KW malignant; human.

XX
OS Homo sapiens.

XX
PN WO9407474-A1.

XX
PD 14-APR-1994.

XX
PF 30-SEP-1993; 93WO-US009338.

XX
PR 30-SEP-1992; 92US-00955012.

XX
PR 25-JUN-1993; 93US-00083590.

XX
PA (UYIA) UNIV YALE.

XX
PI Artavanis-Tsakonas S, Fehon RG, Zagouras P, Blaumueller CM;

XX
DR WPI; 1994-135180/16.

XX
PT Notch protein and nuclear acid compositions - is used for treatment of
PT disorders of cell fate or differentiation esp. breast, colon or cervical
PT cancer.

XX
PS Claim 78; Fig 13; 233pp; English.

XX
CC The invention relates to novel pharmaceutical compositions comprising a
CC therapeutically effective amount of a Notch protein and a
CC pharmaceutically acceptable carrier. The compositions can be used for
CC treatment of disorders of cell fates or differentiation. The therapeutic
CC compositions include Notch proteins and analogues, derivatives and
CC fragments, antibodies, and the nucleic acids encoding the Notch proteins.
CC The compositions are preferably administered to a cancerous condition,
CC e.g. breast, colon or cervical cancer, or to prevent progression from a
CC pre-neoplastic or non-malignant state into a neoplastic or malignant
CC state. This sequence represents a human Notch protein homologue of the
CC invention

XX
SQ Sequence 2471 AA;

Query Match 60.4%; Score 84; DB 2; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKCDXXCNNAACXXDXDC 31
DB 1425 CLSQYCADKARDGVCDACNSHACQWDGDC 1455

RESULT 13
AAY06816
ID AAY06816 standard; protein; 2471 AA.
AC AAY06816;
XX
DT 05-JUL-1999 (first entry)
XX
DE Human Notch2 (humN2) protein sequence.
XX
KW Notch; Nec; Ntm; epidermal growth factor; EGF; repeat domain; cancer;
KW transmembrane domain; cell-fate disorder; proliferative disease;
KW signal transduction; human; Notch2; humN2.
XX
OS Homo sapiens.
XX

Key Location/Qualifiers
FH 25
FT Cleavage-site /note= "potential signal cleavage site"
FT 26..1413
FT Region /note= "EGF repeat region"
FT 1425..1537
FT Region /note= "Lin-12/notch repeats"
FT 1678..1699
FT Domain /note= "transmembrane domain"
FT 1822..2035
FT Region /note= "Ankyrin repeats"
FT 2130..2133
FT Region /note= "nuclear localisation signal (NLS) sequence"
FT 2149..2152
FT Modified-site /note= "putative CKII phosphorylation site"
FT 2156..2159
FT Modified-site /note= "putative cdc2 phosphorylation site"
FT 2166..2169
FT Modified-site /note= "putative cdc2 phosphorylation site"
FT 2172..2174
FT Region /note= "nuclear localisation signal (BNTS) sequence"
FT 2385..2445
FT Region /note= "PEST-containing region"
FT

PN WO9904746-A2.
XX
XX
PD 04-FEB-1999.
XX
XX
PF 23-JUL-1998; 98WO-US015333.
XX
XX
PR 23-JUL-1997; 97US-00899232.
XX
XX
PA (UYVA) UNIV YALE.
XX
PI Artavanis-Tsakonas S, Qi H, Rand MD;
XX
DR WPI; 1999-153290/13.
XX
XX
PT Measuring and detecting activation of Notch protein in cells from cell
PT surface expression - useful for, e.g. identifying modulators of
PT activation, potentially useful for treating cancer.
XX
XX
PS Disclosure; Fig 2A-D; 94pp; English.
XX
CC The invention relates to methods of measuring activation of Notch in a
CC cell. Activated notch can be detected from: (i) expression of: (a) Notch
CC on the cell surface, or (b) one or both of the Notch cleavage products
CC Nec and Ntm; (ii) presence of Notch fragments, i.e. one or both of an N-

terminal fragment (F1) terminating between the epidermal growth factor -
CC like (EGF) repeat domain and the transmembrane domain or a C-terminal
CC fragment (F2) that has its N-terminus between the specified domains, or
CC fragments of molecular weights about 270, 200, 170, 140, 110, 100, 90 and
CC 85 kDa; or (iii) presence of a Notch heterodimer containing a linkage
CC sensitive to reducing agents. The method is used to study or manipulate
CC differentiation processes and to screen/diagnose cell-fate disorders
CC (particularly cancer or other proliferative diseases involving abnormal
CC Notch activation). Modulators of Notch activation are useful for studying
CC or manipulating differentiation and possibly for treating cancer. The F1-
CC F2 heterodimer is the active form of Notch that mediates signal
CC transduction and binds ligands such as Delta and Serrate. F1 and F2 are
CC generated by proteolysis of full-length Notch in the trans-Golgi. The
CC present sequence represents the human Notch2 (humN2) sequence
XX

SQ Sequence 2471 AA;

Query Match 60.4%; Score 84; DB 2; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKCDXXCNNAACXXDXDC 31
DB 1425 CLSQYCADKARDGVCDACNSHACQWDGDC 1455

RESULT 14
AAG79774
ID AAG79774 standard; protein; 2471 AA.
XX
XX
AC AAG79774;
XX
XX
DT 01-APR-2003 (first entry)
XX
DE Human Notch 2.
XX
XX
KW Penetratin; antennapedia; homeodomain; intracellular delivery; anaemia;
KW transport protein; Notch; signalling modulation; rheumatoid arthritis;
KW T-cell mediated disease; asthma; allergy; graft rejection; autoimmunity;
KW tumour; Plasmodium; infection; HIV; Pseudomonas; hepatitis C; measles;
KW multiple sclerosis; diabetes; progressive bulbar palsy; neutropenia;
KW organ; transplantation; bone marrow; cell fate; small cell lung cancer;
KW cancer; kidney; uterus; prostate; bladder; ovary; colon; breast;
KW nervous system; lesion; traumatic; ischaemic; malignant; infectious; HIV;
KW degenerative; demyelinated; diabetes; Parkinson's disease;
KW Charcot-Marie-Tooth disease; liver cirrhosis; Alzheimer's disease;
KW hypertrophic scar formation; psoriasis; organ regeneration;
KW tissue engineering.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200296952-A2.
XX
XX
PD 05-DEC-2002.
XX
XX
PF 24-MAY-2002; 2002WO-GB002438.
XX
XX
PR 25-MAY-2001; 2001GB-00012818.
XX
XX
PA (LORA-) LORANTIS LTD.
XX
PI Solari RCE, Champion BR, Ward GA;
XX
XX
DR WPI; 2003-156790/15.
XX
XX
PT Conjugate for delivering a protein for Notch signal modulation into cells
PT in vitro/in vivo, and in treating T-cell mediated disease, comprises a
PT transport protein sequence and sequence for Notch signaling modulation.
XX
XX
PS Disclosure; Fig 6; 80pp; English.
XX
XX
CC The sequences given in AAG79773-74 represent human Notch 1 and Notch 2
CC which may be used in the conjugate of the invention. The conjugate

CC comprises first and second sequences, where the first sequence comprises
CC a transport protein or a polynucleotide coding for a transport protein
CC and the second sequence comprises a polypeptide or polynucleotide for
CC Notch signalling modulation. The conjugate is useful for transporting a
CC protein for Notch signalling modulation or a polynucleotide sequence
CC encoding it into a cell. Pharmaceuticals comprising the conjugate and the
CC invention are useful in the treatment of a T-cell mediated disease and
CC for the preparation of a medicament for the preventing and/or treatment
CC of disease or infection, especially T-cell mediated disease. T-cell
CC mediated diseases include asthma, allergy, graft rejection, autoimmune
CC tumour induced aberrations to the T cell system and infectious diseases
CC such as caused by Plasmodium sp., HIV, Pseudomonas, hepatitis C, measles,
CC multiple sclerosis, rheumatoid arthritis, diabetes, organ transplantation
CC or bone marrow transplantation. The conjugate is also useful for altering
CC the fate of a cell, tissue or organ type by altering Notch pathway
CC function in the cell, in the treatment of malignant and pre-neoplastic
CC disorders e.g. small cell lung cancer and cancers of the kidney, uterus,
CC prostate, bladder, ovary, colon and breast. It is further useful for
CC treating nervous system disorders including neurological lesions
CC including traumatic lesions resulting from physical injuries, ischaemic
CC lesions, malignant lesions, infectious lesions caused by HIV,
CC degenerative lesions and diseases and demyelinated lesions. Diseases
CC treatable include diabetes, Parkinson's, Alzheimer's disease, progressive
CC bulbar palsy and Charcot-Marie-Tooth disease. The conjugate is also
CC useful for promoting tissue regeneration and repair, treating diseases
CC associated with defective tissue repair and regeneration for e.g.
CC cirrhosis of the liver, hypertrophic scar formation and psoriasis, and
CC also for treating neutropenia, or anaemia and in techniques of organ
CC regeneration and tissue engineering. The conjugate has improved efficacy
CC compared to conventional treatments, improved cellular uptake of the
CC therapeutic agent, improved water solubility, reduction of side effects
CC and cellular bioavailability and decreased occurrence of drug resistance
XX
SQ Sequence 2471 AA;

Query Match 60.4%; Score 84; DB 6; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDACNSHACQWDGDC 1455

RESULT 15

ABP72572
ID ABP72572 standard; protein; 2471 AA.
XX
AC ABP72572;
XX
DT 29-MAY-2003 (first entry)
XX
DE Human Notch 2.
XX
KW Human; Notch 2; signal transduction; modulator; cyostatic;
KW antiasthmatic; antiallergic; antirheumatic; antiarthritic; antidiabetic;
KW antithyroid; neuroprotective; ophthalmological; antiinflammatory;
KW hepatotropic; dermatological; immunosuppressive; antibacterial; virucide;
KW antiarteriosclerotic; cardiant; antiarteriosclerotic; vasotropic;
KW antidiabetic; nephrotropic; antiparkinsonian; nootropic; gynaecological;
KW protozoacide.
XX
OS Homo sapiens.
XX
PN WO2003012441-A1.
XX
PD 13-FEB-2003.
XX
PF 25-JUL-2002; 2002WO-GB003397.
XX
PR 25-JUL-2001; 2001GB-00018153.
PR 05-APR-2002; 2002GB-00007930.
PR 28-MAY-2002; 2002GB-00012282.

PR 28-MAY-2002; 2002GB-00012283.
XX
XX (LORA-) LORANTIS LTD.
PA
XX
PI Bodmer MW, Briend ECP, Champion BR, McKenzie GJ, Tugal T;
PI Ward GA, Young LB;
XX
DR WPI: 2003-248208/24.

XX
PT Detecting modulators of notch or immune signaling useful for treating
PT cancer, asthma, allergy, lupus erythematosus, diabetes, by monitoring
PT Notch signaling in a cell of the immune system in the presence of
PT modulator.

PS Disclosure; Fig 36; 184pp; English.

XX
XX The present sequence is the protein sequence of human Notch 2. The
CC invention provides a method for detecting modulators of Notch signalling.
CC This involves monitoring the effect of a candidate modulator on Notch
CC signalling in a cell of the immune system, such as a T cell, a B cell or
CC an antigen presenting cell, e.g. by monitoring the level of expression of
CC a target gene such as an endogenous target gene of the Notch signalling
CC pathway or a reporter gene. The modulator may comprise a Notch
CC intracellular domain (see also ABP72552) or a polynucleotide encoding it.
CC The method detects modulators of Notch signalling by monitoring the
CC effect of a candidate modulator on Notch signalling in a cell of the
CC immune system, such as a T cell, a B cell or an antigen presenting cell.
CC The method is useful for preparing a medicament for treating a disease or
CC condition of, or related to, the immune system, such as T-cell, B-cell or
CC APC mediated disease (all claimed), including cancer, acute and chronic
CC immune and autoimmune pathologies, infectious diseases, inflammatory
CC diseases, neurodegenerative diseases, alcohol-induced hepatitis, other
CC diseases related to angiogenesis, cardiovascular conditions and pulmonary
CC diseases
XX
SQ Sequence 2471 AA;

Query Match 60.4%; Score 84; DB 6; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDACNSHACQWDGDC 1455

Search completed: May 19, 2004, 15:57:48
Job time : 5.0076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:55:54 ; Search time 1.29658 Seconds
(Without alignments)
1234.330 Million cell updates/sec

Title: US-09-903-199-4
Perfect score: 139
Sequence: 1 CDXXXCXKXGXGNGXCDXXCNMAACXXDXDC 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfil1est1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	105	75.5	1139	1	US-08-537-210A-4	Sequence 4, Appli
2	105	75.5	1139	3	US-09-113-825-4	Sequence 4, Appli
3	105	75.5	2703	1	US-08-185-432-19	Sequence 19, Appli
4	105	75.5	2703	4	US-08-899-232-4	Sequence 4, Appli
5	84	60.4	1015	1	US-08-537-210A-1	Sequence 1, Appli
6	84	60.4	1015	3	US-09-113-825-1	Sequence 1, Appli
7	84	60.4	1068	3	US-08-537-210A-2	Sequence 2, Appli
8	84	60.4	1068	3	US-09-113-825-2	Sequence 2, Appli
9	84	60.4	2471	1	US-08-185-432-16	Sequence 16, Appli
10	84	60.4	2471	1	US-08-083-590A-19	Sequence 19, Appli
11	84	60.4	2471	3	US-08-532-384-19	Sequence 19, Appli
12	84	60.4	2471	4	US-08-899-232-1	Sequence 1, Appli
13	84	60.4	2556	1	US-08-185-432-17	Sequence 17, Appli
14	84	60.4	2556	1	US-08-083-590A-20	Sequence 20, Appli
15	84	60.4	2556	3	US-08-532-384-20	Sequence 20, Appli
16	84	60.4	2556	4	US-08-899-232-2	Sequence 2, Appli
17	83	59.7	2321	4	US-09-230-652-2	Sequence 2, Appli
18	80	57.6	1064	1	US-08-537-210A-3	Sequence 3, Appli
19	80	57.6	1064	3	US-09-113-825-3	Sequence 3, Appli
20	80	57.6	2523	1	US-08-185-432-18	Sequence 18, Appli
21	80	57.6	2523	4	US-08-899-232-3	Sequence 3, Appli
22	79	56.8	1964	4	US-09-467-997-1	Sequence 1, Appli
23	78	56.1	928	4	US-09-635-872A-1	Sequence 1, Appli
24	78	56.1	928	4	US-09-636-077A-1	Sequence 1, Appli
25	78	56.1	928	4	US-09-636-060C-1	Sequence 1, Appli
26	78	56.1	928	4	US-09-986-552-1	Sequence 1, Appli
27	77	55.4	1078	1	US-08-264-534-32	Sequence 32, Appli

28	77	55.4	1078	1	US-08-083-590A-11	Sequence 11, Appli
29	77	55.4	1078	1	US-08-465-500-32	Sequence 32, Appli
30	77	55.4	1078	2	US-08-346-128-32	Sequence 32, Appli
31	77	55.4	1078	3	US-08-532-384-11	Sequence 11, Appli
32	77	55.4	1078	3	US-08-893-828-32	Sequence 32, Appli
33	73	52.5	502	4	US-09-635-872A-13	Sequence 13, Appli
34	73	52.5	502	4	US-09-636-077A-13	Sequence 13, Appli
35	73	52.5	502	4	US-09-636-060C-13	Sequence 13, Appli
36	73	52.5	502	4	US-09-986-552-13	Sequence 13, Appli
37	64	46.0	1791	4	US-09-827-998-3	Sequence 3, Appli
38	62	44.6	908	4	US-09-635-872A-15	Sequence 15, Appli
39	62	44.6	908	4	US-09-636-077A-15	Sequence 15, Appli
40	62	44.6	908	4	US-09-636-060C-15	Sequence 15, Appli
41	62	44.6	908	4	US-09-986-552-15	Sequence 15, Appli
42	59	42.4	1385	4	US-09-827-998-16	Sequence 16, Appli
43	57	41.0	1770	4	US-09-827-998-10	Sequence 10, Appli
44	48.5	34.9	233	4	US-09-216-393B-110	Sequence 110, App
45	48.5	34.9	616	4	US-09-608-790-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-537-210A-4
Sequence 4, Application US/08537210A
Patent No. 5780300

GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1139 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Dros N
LOCATION: 1189...2327
OTHER INFORMATION: Highly conserved ankyrin repeat
region of No. 57803000ch
US-08-537-210A-4

Query Match 75.5%; Score 105; DB 1; Length 1139;

QY 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDCACNSHACQWDGDC 1455

RESULT 10

US-08-083-590A-19
Sequence 19, Application US/08083590A

Patent No. 5786158

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

TITLE OF INVENTION: Nucleic Acids

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083,590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2471 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-083-590A-19

Query Match 60.4%; Score 84; DB 1; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.003;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDCACNSHACQWDGDC 1455

RESULT 11
US-08-532-384-19
Sequence 19, Application US/08532384

Patent No. 6083904

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

TITLE OF INVENTION: Nucleic Acids

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/532,384

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/083,590

FILING DATE: 25-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2471 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-532-384-19
Query Match 60.4%; Score 84; DB 3; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.003;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDCACNSHACQWDGDC 1455

RESULT 12

US-08-899-232-1

Sequence 1, Application US/08899232

Patent No. 6436650

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon

TITLE OF INVENTION: Qi, Huilin

TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON

FILE REFERENCE: 7326-046

CURRENT APPLICATION NUMBER: US/08/899,232

CURRENT FILING DATE: 1997-07-23

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 2471

TYPE: PRT

ORGANISM: Homo sapiens

US-08-899-232-1

Query Match 60.4%; Score 84; DB 4; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.003;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGVCDCACNSHACQWDGDC 1455

RESULT 13

US-08-185-432-17

Sequence 17, Application US/08185432

Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-17

Query Match 60.4%; Score 84; DB 1; Length 2556;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGXGXCDDXXCENNAACXXDXDC 31
Db 1450 CELPEQEDAGNKVCSLQCNHACGWDGDC 1480

RESULT 14
US-08-083-590A-20
Sequence 20, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,590A
FILING DATE: 25-JUN-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match 60.4%; Score 84; DB 1; Length 2556;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGXGXCDDXXCENNAACXXDXDC 31
Db 1450 CELPEQEDAGNKVCSLQCNHACGWDGDC 1480

RESULT 15
US-08-532-384-20
Sequence 20, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-20

Query Match	60.4%	Score 84	DB 3	Length 2556
Best Local Similarity	45.2%	Pred. No. 0.0031		
Matches 14, Conservative		1, Mismatches 16	Indels 0	Gaps 0

QY 1 CDXXXCXXKXGNGXCDDXXCNNAAACXXDXDC 31
| : | | | | | | | | | | | | | | | | | | | |
DB 1450 CELPECGEDAGNKVKYCSLQCNNHACGWDGDC 1480

Search completed: May 19, 2004, 16:01:21
Job time : 1.29658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:59:55 ; Search time 3.06464 Seconds

(without alignments)
2814.722 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139

Sequence: 1 CDXXXCXKXGNGXCDXXCNMAACXXDXDC 31

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	115	82.7	31	9	US-09-903-248-4	Sequence 4, Appli
2	115	82.7	31	9	US-09-859-604-4	Sequence 4, Appli
3	115	82.7	31	9	US-09-903-063-4	Sequence 4, Appli
4	115	82.7	31	9	US-09-903-216-4	Sequence 4, Appli
5	115	82.7	31	9	US-09-903-199-4	Sequence 4, Appli
6	115	82.7	31	9	US-09-903-023-4	Sequence 4, Appli
7	115	82.7	31	10	US-09-436-184-4	Sequence 4, Appli
8	105	75.5	1139	12	US-10-419-026-4	Sequence 4, Appli
9	100	71.9	39	12	US-10-072-012-825	Sequence 825, App
10	100	71.9	39	15	US-10-190-115-83	Sequence 83, Appli
11	100	71.9	39	15	US-10-190-115-84	Sequence 84, Appli
12	100	71.9	39	15	US-10-369-072-83	Sequence 83, Appli
13	100	71.9	39	15	US-10-369-072-84	Sequence 84, Appli
14	92	66.2	1429	15	US-10-369-493-5819	Sequence 5819, App
15	85	61.2	2531	12	US-10-072-012-470	Sequence 470, App

16	85	61.2	2531	12	US-10-072-012-471	Sequence 471, App
17	85	61.2	2531	15	US-10-190-115-29	Sequence 29, Appli
18	85	61.2	2531	15	US-10-369-072-29	Sequence 29, Appli
19	84	60.4	874	15	US-10-294-006-34	Sequence 34, Appli
20	84	60.4	1015	12	US-10-419-026-1	Sequence 1, Appli
21	84	60.4	1068	12	US-10-419-026-2	Sequence 2, Appli
22	84	60.4	2444	9	US-09-944-849-2	Sequence 2, Appli
23	84	60.4	2444	12	US-10-072-012-469	Sequence 469, App
24	84	60.4	2469	15	US-10-190-115-2	Sequence 2, Appli
25	84	60.4	2469	15	US-10-369-072-2	Sequence 2, Appli
26	84	60.4	2471	15	US-10-190-115-27	Sequence 27, Appli
27	84	60.4	2471	15	US-10-369-072-27	Sequence 27, Appli
28	84	60.4	2555	12	US-10-072-012-468	Sequence 468, App
29	84	60.4	2556	12	US-10-072-012-134	Sequence 134, App
30	84	60.4	2556	12	US-10-072-012-467	Sequence 467, App
31	84	60.4	2556	15	US-10-294-006-12	Sequence 12, Appli
32	83	59.7	2321	14	US-10-356-625-2	Sequence 2, Appli
33	82	59.0	2317	15	US-10-190-115-26	Sequence 26, Appli
34	82	59.0	2317	15	US-10-369-072-26	Sequence 26, Appli
35	80	57.6	1064	12	US-10-419-026-3	Sequence 3, Appli
36	80	57.6	2524	15	US-10-190-115-25	Sequence 25, Appli
37	80	57.6	2524	15	US-10-369-072-25	Sequence 25, Appli
38	79	56.8	1295	15	US-10-369-493-5820	Sequence 5820, App
39	79	56.8	1295	15	US-10-369-493-5821	Sequence 5821, App
40	78	56.1	846	15	US-10-094-749-2392	Sequence 2392, App
41	78	56.1	847	15	US-10-120-801-88	Sequence 88, Appli
42	78	56.1	928	9	US-09-895-072-1	Sequence 1, Appli
43	78	56.1	928	9	US-09-986-552-1	Sequence 1, Appli
44	78	56.1	928	14	US-10-023-888-4	Sequence 4, Appli
45	78	56.1	928	14	US-10-023-889-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-903-248-4
; Sequence 4, Application US/09903248
; Patent No. US20020102263A1
; GENERAL INFORMATION:
; APPLICANT: Wande, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIVS
; CURRENT APPLICATION NUMBER: US/09/903, 248
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; NAME/KEY: VARIANT
; LOCATION: (3)..(5)
; OTHER INFORMATION: wherein any xaa may be any amino acid
; NAME/KEY: VARIANT
; LOCATION: (6)..(7)
; OTHER INFORMATION: wherein xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (10)
; OTHER INFORMATION: wherein xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (14)
; OTHER INFORMATION: wherein xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (17)..(18)


```
; NAME/KEY: VARIANT
; LOCATION: (25)..(26)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (29)
; OTHER INFORMATION: wherein Xaa is any amino acid.
US-09-903-248-4
```

```
Query Match      82.7%; Score 115; DB 9; length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
Db
```

```
RESULT 2
US-09-859-604-4
; Sequence 4, Application US/09859604
; Patent No. US20020110559A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M
; APPLICANT: Deutch, Alan H
; APPLICANT: Chahbari, Hossein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
```

```
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; OTHER INFORMATION: cysteine-rich repeat
; NAME/KEY: VARIANT
; LOCATION: (2)..(5)
; OTHER INFORMATION: wherein any Xaa may be any amino acid
; NAME/KEY: VARIANT
; LOCATION: (7)..(8)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (10)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (14)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (17)..(18)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (25)..(26)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (29)
; OTHER INFORMATION: wherein Xaa is any amino acid.
US-09-859-604-4
```

```
Query Match      82.7%; Score 115; DB 9; length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
Db
```

```
RESULT 3
US-09-903-063-4
; Sequence 4, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
```

```
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; OTHER INFORMATION: cysteine-rich repeat
; NAME/KEY: VARIANT
; LOCATION: (3)..(5)
; OTHER INFORMATION: wherein any Xaa may be any amino acid
; NAME/KEY: VARIANT
; LOCATION: (6)..(7)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (10)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (14)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (17)..(18)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (25)..(26)
; OTHER INFORMATION: wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (29)
; OTHER INFORMATION: wherein Xaa is any amino acid.
US-09-903-063-4
```

```
Query Match      82.7%; Score 115; DB 9; length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
      1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
Db
```

```
RESULT 4
US-09-903-216-4
; Sequence 4, Application US/09903216
; Patent No. US20020114811A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV2
; CURRENT APPLICATION NUMBER: US/09/903,216
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
```

```
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: EGF-like
NAME/KEY: VARIANT
LOCATION: (3)..(5)
OTHER INFORMATION: Wherein Xaa may be any amino acid
NAME/KEY: VARIANT
LOCATION: (6)..(7)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (10)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (14)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (17)..(18)
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-216-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
```

```
RESULT 5
US-09-903-199-4
; Sequence 4, Application US/09903199
; Patent No. US20020122802A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV4
; CURRENT APPLICATION NUMBER: US/09/903,199
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; OTHER INFORMATION: cysteine-rich repeat
; NAME/KEY: VARIANT
; LOCATION: (3)..(5)
; OTHER INFORMATION: Wherein Xaa may be any amino acid
; NAME/KEY: VARIANT
; LOCATION: (6)..(7)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (10)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (14)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
```

```
NAME/KEY: VARIANT
LOCATION: (17)..(18)
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-199-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
```

```
RESULT 6
US-09-903-023-4
; Sequence 4, Application US/09903023
; Patent No. US20020146421A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV1
; CURRENT APPLICATION NUMBER: US/09/903,023
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; OTHER INFORMATION: cysteine-rich repeat
; NAME/KEY: VARIANT
; LOCATION: (3)..(5)
; OTHER INFORMATION: Wherein Xaa may be any amino acid
; NAME/KEY: VARIANT
; LOCATION: (6)..(7)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (10)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (14)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (17)..(18)
; NAME/KEY: VARIANT
; LOCATION: (25)..(26)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; NAME/KEY: VARIANT
; LOCATION: (29)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-023-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db      1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
```

```

RESULT 7
US-09-436-184-4
; Sequence 4, Application US/09436184
; Publication No. US20030031670A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: R.I. Hosp. - Malignant Neoplasms
; CURRENT APPLICATION NUMBER: US/09/436,184
; CURRENT FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: EGF-like
; OTHER INFORMATION: cysteine-rich repeat
; OTHER INFORMATION: wherein any Xaa may be any amino acid
US-09-436-184-4

```

Query Match	82.7%;	Score 115;	DB 10;	Length 31;
Best Local Similarity	100.0%;	Pred. No. 5.4e-10;		
Matches 31;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 CDXXXXCXXKXGNGXCDXXCINNAACXXDXDC 31
Dy 1 CDXXXXCXXKXGNGXCDXXCINNAACXXDXDC 31
Db 1 CDXXXXCXXKXGNGXCDXXCINNAACXXDXDC 31

```

RESULT 8
US-10-419-026-4
; Sequence 4, Application US/10419026
; Publication No. US20040058443A1
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Teakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsumo, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY DIFFERENTIATED CELLS USING THE
; TITLE OF INVENTION: NOTCH PATHWAY
; FILE REFERENCE: 10910-096
; CURRENT APPLICATION NUMBER: US/10/419,026
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 08/537,210
; PRIOR FILING DATE: 1995-09-29
; PRIOR APPLICATION NUMBER: 09/113,824
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-419-026-4

```

Query Match	75.5%;	Score 105;	DB 12;	Length 1139;
Best Local Similarity	54.8%;	Pred. No. 4.2e-07;		
Matches 17; Conservative		0; Mismatches 14;	Indels 0;	Gaps 0;

QY 1 CDXXXCKKKGNXCDYXCNNNAACXXDXDC 31
Db 294 CDKRGCTEKQNGICDSDCNTTYACNFDGND 324

RESULT 9
US-10-072-012-825
; Sequence 825, Application US/10072012
; Publication No. US20040033493A1

```

: GENERAL INFORMATION:
: APPLICANT: Tchernev, Velizar
: APPLICANT: Spytek, Kimberly
: APPLICANT: Zerhusen, Bryan
: APPLICANT: Patcurajan, Meera
: APPLICANT: Shinkets, Richard
: APPLICANT: Li, Li
: APPLICANT: Gangolli, Esba
: APPLICANT: Padigaru, Muralidhara
: APPLICANT: Anderson, David W.
: APPLICANT: Rastelli, Luca
: APPLICANT: Miller, Charles E.
: APPLICANT: Gerlach, Valerie
: APPLICANT: Taupier Jr, Raymond J.
: APPLICANT: Gusev, Vladimir Y.
: APPLICANT: Colman, Steven D.
: APPLICANT: Wolen, Adam R.
: APPLICANT: Pena, Carol E. A
: APPLICANT: Furtak, Katarzyna
: APPLICANT: Grose, William M.
: APPLICANT: Alsbrook II, John P.
: APPLICANT: Lepley, Denise M.
: APPLICANT: Rieger, Daniel K.
: APPLICANT: Burgess, Catherine E.
: TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-258
: CURRENT APPLICATION NUMBER: US/10/072,012
: CURRENT FILING DATE: 2002-01-31
: PRIOR APPLICATION NUMBER: 60/265,102
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: 60/265,514
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,517
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,412
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,395
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/266,406
: PRIOR FILING DATE: 2001-02-02
: PRIOR APPLICATION NUMBER: 60/266,767
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: 60/267,057
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/266,975
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/267,459
: PRIOR FILING DATE: 2001-02-08
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 1391
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 825
: LENGTH: 39
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Domain found
: OTHER INFORMATION: in No. US20040033493A1ch and Lin-12
US-10-072-012-825

```

Query Match	71.9%;	Score 100;	DB 12;	Length 39;
Best Local Similarity	51.6%;	Pred. NO. 1e-07;		
Matches	16;	Conservative	2;	Mismatches 13;
			Indels	0;
			Gaps	0;

Q7 1 CDXXXXCXXKXGNGXCDXXCINNAACXXDXDC 31
 | : | : | : | : | : | : | : |
D6 8 CEDAQCMDKFEGDVCDEECNNAECLWDGDC 38

RESULT 10
US-10-190-115-83
; Sequence 83, Application US/10190115
; Publication No. US20030207394A1

```

; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John P. II
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Grosse, William M.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Ji, Weizhen
; APPLICANT: Lepley, Denise M.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Mezick, Amanda J.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: Shen, Lei
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Szekeres, Edward S. Jr.
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Zernhusen, Bryan D.
; APPLICANT: Voss, Edward Z.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-050 CIP
; CURRENT APPLICATION NUMBER: US/10/190,115
; PRIOR FILING DATE: 2003-02-10
; PRIOR APPLICATION NUMBER: 60/303,168
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/368,996
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 60/386,816
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585,
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2001-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: Curaseqqlist version 0.1
; SEQ ID NO 83
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-190-115-83

Query Match      71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0,

QY      1 CDXXXKXXKXGXCDXXCXXNNAACXXDXDC 31
      1: | | | | | | | | | | | | | | |
Db      8 CEDAQCWDFGDCVCDCECNAECCLWDGDC 38

RESULT 11
US-10-190-115-84
; Sequence 84, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John P. II
; APPLICANT: Boldog, Ferenc L.

```

```

: APPLICANT: Burgess, Catherine E.
: APPLICANT: Casman, Stacie J.
: APPLICANT: Grosse, William M.
: APPLICANT: Gusev, Vladimir Y.
: APPLICANT: Ji, Weizhen
: APPLICANT: Lepley, Denise M.
: APPLICANT: Liu, Xiaohong
: APPLICANT: Meziack, Amanda J.
: APPLICANT: Padigaru, Muralidhara
: APPLICANT: Patturajan, Meera
: APPLICANT: Rastelli, Luca
: APPLICANT: Shen, Lei
: APPLICANT: Shenoy, Suresh G.
: APPLICANT: Shinkets, Richard A.
: APPLICANT: Spaderna, Steven K.
: APPLICANT: Spytek, Kimberly A.
: APPLICANT: Szekeres, Edward S. Jr.
: APPLICANT: Taupier, Raymond J. Jr.
: APPLICANT: Tchernev, Velizar T.
: APPLICANT: Zerhusen, Bryan D.
: APPLICANT: Voss, Edward Z.
: TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 21402-050 CIP
: CURRENT APPLICATION NUMBER: US/10/190,115
: CURRENT FILING DATE: 2003-02-10
: PRIOR APPLICATION NUMBER: 60/303,168
: PRIOR FILING DATE: 2001-07-05
: PRIOR APPLICATION NUMBER: 60/368,996
: PRIOR FILING DATE: 2002-04-01
: PRIOR APPLICATION NUMBER: 60/386,816
: PRIOR FILING DATE: 2002-06-07
: PRIOR APPLICATION NUMBER: 60/215,854
: PRIOR FILING DATE: 2000-07-03
: PRIOR APPLICATION NUMBER: 60/215,856
: PRIOR FILING DATE: 2000-07-03
: PRIOR APPLICATION NUMBER: 60/215,902
: PRIOR FILING DATE: 2000-07-03
: PRIOR APPLICATION NUMBER: 60/216,585,
: PRIOR FILING DATE: 2000-07-07
: PRIOR APPLICATION NUMBER: 60/216,586
: PRIOR FILING DATE: 2001-07-07
: PRIOR APPLICATION NUMBER: 60/216,722
: PRIOR FILING DATE: 2000-07-07
: PRIOR APPLICATION NUMBER: 60/218,622
: PRIOR FILING DATE: 2000-07-17
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 136
: SOFTWARE: Curaseq1ist version 0.1
: SEQ ID NO 84
: LENGTH: 39
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-190-115-84

Query Match      71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY      1 CDXXXKXXKXGXGXCXXXCNNAACXXDXDC 31
      | : | : | : | : | : | : | : | : |
DB      8 CEDAQCMDFGDCVCECCNNAECLMDGGDC 38

RESULT 12
US-10-369-072-83
: Sequence 83, Application US/10369072
: Publication No. US20040014081A1
: GENERAL INFORMATION:
: APPLICANT: Alsobrook II, John P
: APPLICANT: Spaderna, Stephen K
: APPLICANT: Tchernev, Velizar
: APPLICANT: Liu, Xiaohong
: APPLICANT: Shenoy, Suresh

```

```

; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
; APPLICANT: Szerkeres, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CON2
; CURRENT APPLICATION NUMBER: US/10/369,072
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 83
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Domain search
; OTHER INFORMATION: result
US-10-369-072-83

Query Match          71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY      1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31
      | : | | | | | | | | | | | | | | |
Db      8 CEDAQCMWDFKFGDVCDEECNNAECIWDGGDC 38

RESULT 13
US-10-369-072-84
; Sequence 84, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Spaderna, Stephen K
; APPLICANT: Tchernev, Velizar
; APPLICANT: Liu, Xiaohong
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
```

```

; APPLICANT: Szerkeres, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CON2
; CURRENT APPLICATION NUMBER: US/10/369,072
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 84
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Domain search
; OTHER INFORMATION: result
US-10-369-072-84

Query Match          71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY      1 CDXXXCKXKXGNGXCDXXCNNAACXXDXDC 31
      | : | | | | | | | | | | | | | | |
Db      8 CEDAQCMWDFKFGDVCDEECNNAECIWDGGDC 38

RESULT 14
US-10-369-493-5819
; Sequence 5819, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5819
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
```


US-10-369-493-5819

Query Match 66.2%; Score 92; DB 15; Length 1429;
Best Local Similarity 45.2%; Pred. No. 4e-05;
Matches 14; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 638 CEKRKCSERANDGNCADACNYAACRFDGDC 668

RESULT 15
US-10-072-012-470

; Sequence 470, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:

; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Eshta
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258

; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentn Ver. 2.1

; SEQ ID NO 470
; LENGTH: 2531
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-470

Query Match 61.2%; Score 85; DB 12; Length 2531;

Best Local Similarity 45.2%; Pred. No. 0.00071;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1449 CELPECEDEAGNKKVCNLCNNHACGWDGDC 1479

Search completed: May 19, 2004, 16:09:34
Job time : 3.06464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:54:54 ; Search time 1.29658 seconds

(without alignments)
2299.852 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139

Sequence: 1 CDXXXXXXKXGNGXCDXXCNNAACXXDXDC 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*
1: pirl:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	75.5	2703	1 A24420	notch protein - fr
2	97	69.8	2437	2 S42612	transmembrane prot
3	92	66.2	1429	2 S06434	homeotic protein I
4	91	65.5	2531	2 T31070	notch homolog - se
5	88	63.3	1203	2 A49175	Notch B protein -
6	85	61.2	387	2 B49175	Notch A protein -
7	85	61.2	861	2 A48825	Notch homolog Motc
8	85	61.2	2531	2 S18188	notch protein homo
9	85	61.2	2531	2 A46019	notch-1 protein -
10	84	60.4	2471	2 A49128	cell-fate determin
11	84	60.4	2555	2 A40043	notch protein homo
12	83	59.7	2321	2 S78549	notch3 protein - h
13	82	59.0	2318	2 S45306	notch 3 protein -
14	80	57.6	2524	2 A35844	Xotch protein - Af
15	79	56.8	1295	2 A32901	glp1 protein precu
16	79	56.8	1964	2 T09059	notch4 - mouse
17	77	55.4	2352	2 T30201	Notch homolog prot
18	61	43.9	67	2 T11547	metallothionein -
19	60	43.2	1627	2 S65464	pregnancy-associat
20	52	37.4	2533	2 T28675	alpha-51D immobili
21	52	37.4	2533	2 T28674	alpha-51D-immobil
22	50	36.0	65	2 A38739	metallothionein -
23	49	35.3	2543	2 T31687	surface antigen - P
24	48.5	34.9	616	2 A55796	ecarin precursor -
25	47.5	34.2	2395	1 S50820	surface protein ty
26	46.5	33.5	419	2 A59414	metalloproteinase
27	46.5	33.5	571	2 S24789	jararhagin C precu
28	46.5	33.5	609	2 S55270	catrocollastatin p
29	46	33.1	575	1 THHUB	thrombomodulin pre

30	45	32.4	64	2 A33825	metallothionein A
31	45	32.4	71	2 A59412	KGD-bearing plate
32	45	32.4	72	2 B43019	platelet aggregati
33	45	32.4	72	2 D43019	platelet aggregati
34	45	32.4	73	2 A43019	platelet aggregati
35	45	32.4	73	2 C43019	platelet aggregati
36	45	32.4	73	2 B40003	platelet aggregati
37	45	32.4	73	2 A40003	platelet aggregati
38	45	32.4	2706	2 T28155	variant-specific s
39	45	32.4	2907	2 A57278	fibrillin-2 precu
40	45	32.4	2918	2 A54105	fibrillin-2 precu
41	44.5	32.0	64	2 A25775	metallothionein A
42	44.5	32.0	313	2 S44208	extracellular matr
43	44.5	32.0	788	2 A37057	integrin beta-6 ch
44	44	31.7	186	2 A28401	agglutinin isolect
45	44	31.7	673	2 A48089	growth arrest-spec

ALIGNMENTS

RESULT 1

A24420

notch protein - fruit fly (Drosophila melanogaster)

N;Alternate names: neurogenic repetitive locus protein

C;Species: Drosophila melanogaster	
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999	
C;Accession: A24420; A24768; S09358; A05267	
R;Kidd, S.; Kelley, M.R.; Young, M.W.	
Mol. Cell. Biol. 6, 3094-3108, 1986	
A;Reference number: A24420; MUID:87064624; PMID:3097517	
A;Accession: A24420	
A;Molecule type: DNA	
A;Residues: 1-2703 <KID>	
A;Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993	
R;Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.	
Cell 43, 567-581, 1985	
A;Reference number: A24768; MUID:86079539; PMID:3935325	
A;Accession: A24768	
A;Molecule type: mRNA	
A;Residues: 1-48, 'I', 50-118, 'R', 120-230, 'I', 232-256, 'N', 258-266, 'A', 268-872, 'R', 874-958	
A;Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 204	
R;Tautz, D.	
Nucleic Acids Res. 17, 6463-6471, 1989	
A;Title: Hypervariability of simple sequences as a general source for polymorphic DNA m	
A;Reference number: S09358; MUID:89385974; PMID:2780284	
A;Accession: S09358	
A;Molecule type: DNA	
A;Residues: 2505-2551, 'QOOQ', 2552-2576, 'E', 2578-2604 <TAU>	
R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.	
Cell 40, 55-62, 1985	
A;Title: opa: a novel family of transcribed repeats shared by the Notch locus and other	
A;Reference number: A05267; MUID:85099329; PMID:2981631	
A;Accession: A05267	
A;Molecule type: DNA	
A;Residues: 2504-2576, 'E', 2578-2611 <WHA2>	
C;Genetics:	
A;Gene: notch; opa	
A;Cross-references: FlyBase:FBgn0004647	
A;Map position: 8.96-9.36	
A;Intons: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3	
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology	
C;Keywords: differentiation; tandem repeat; transmembrane protein	
F;27-43/Domain: transmembrane #status predicted <TMM1>	
F;297-328/Domain: EGF homology <EGX1>	
F;530-561/Domain: EGF homology <EGF1>	
F;568-599/Domain: EGF homology <EGF>	
F;988-1019/Domain: EGF homology <EGX2>	
F;1064-1095/Domain: EGF homology <EGF3>	
F;1187-1218/Domain: EGF homology <EGF3>	
F;1746-1762/Domain: transmembrane #status predicted <TMM2>	
F;1950-1982/Domain: ankyrin repeat homology <AN1>	
F;1983-2015/Domain: ankyrin repeat homology <AN2>	

F;1988-2004/Domain: transmembrane #status predicted <TM3>
F;2017-2049/Domain: ankyrin repeat homology <AN3>
F;2050-2082/Domain: ankyrin repeat homology <AN4>
F;2083-2115/Domain: ankyrin repeat homology <AN5>
F;2538-2568/Region: glutamine-rich
F;2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 75.5%; Score 105; DB 1; Length 2703;
Best Local Similarity 54.8%; Pred. No. 3.1e-07;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1482 CDKRGCTEKQNGICDSDCNTYACNFDGND 1512

RESULT 2

transmembrane protein precursor - zebra fish
C;Species: Brachydanio rerio (zebra fish)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
C;Accession: S42612
R;Bierkamp, C.; Campos-Ortega, J.A.
Mech. Dev. 43, 87-100, 1993

A;Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern
A;Reference number: S42612; MUID:94128602; PMID:8297791
A;Accession: S42612

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2437 <BIE>

A;Cross-references: EMBL:X69088; NID:g433866; PIDN:CAA48831.1; PID:g433867
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;755-786/Domain: EGF homology <EGF1>
F;1023-1054/Domain: EGF homology <EGF2>
F;1185-1216/Domain: EGF homology <EGF2>
F;1915-1947/Domain: ankyrin repeat homology <AN1>
F;1948-1980/Domain: ankyrin repeat homology <AN2>
F;1982-2014/Domain: ankyrin repeat homology <AN3>
F;2015-2047/Domain: ankyrin repeat homology <AN4>
F;2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 69.8%; Score 97; DB 2; Length 2437;
Best Local Similarity 48.4%; Pred. No. 3.9e-06;
Matches 15; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

OY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1447 CETAQCEGRGNAICDTQCNNHACGWDGDC 1477

RESULT 3

homeotic protein lin-12 precursor - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 20-Sep-1999
C;Accession: S06434; A24769
R;Vochem, J.; Weston, K.; Greenwald, I.

Nature 335, 547-550, 1988

A;Title: The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with ove
A;Reference number: S06434; MUID:88334747; PMID:3419531
A;Accession: S06434

A;Molecule type: DNA

A;Residues: 1-1429 <YOC>

A;Cross-references: EMBL:M12069; NID:g156357; PIDN:AAA70191.1; PID:g156358
R;Greenwald, I.

Cell 43, 583-590, 1985

A;Reference number: A24769; MUID:86079540; PMID:3000611

A;Accession: A24769

A;Molecule type: DNA

A;Residues: 173-712 <GRE>

C;Genetics:

A;Introns: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3

C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C;Keywords: glycoprotein; transmembrane protein
F;254-284/Domain: EGF homology <EGF1>
F;507-540/Domain: EGF homology <EGF2>
F;547-578/Domain: EGF homology <EGF2>
F;909-931/Domain: transmembrane #status predicted <TM3>
F;1093-1125/Domain: ankyrin repeat homology <AN1>
F;1206-1238/Domain: ankyrin repeat homology <AN2>
F;1240-1272/Domain: ankyrin repeat homology <AN3>

Query Match 66.2%; Score 92; DB 2; Length 1429;
Best Local Similarity 45.2%; Pred. No. 1.4e-05;
Matches 14; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

OY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 638 CERKCSERANDGNCADCNVYACKFDGDC 668

RESULT 4

T31070

notch homolog - sea urchin (Lytechinus variegatus)

C;Species: Lytechinus variegatus (variegated urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C;Accession: T31070

R;Sherwood, D.R.; McClay, D.R.
Development 124, 3363-3374, 1997

A;Title: Identification and localization of a sea urchin Notch homologue: insights into

A;Reference number: Z20966; MUID:97454256; PMID:9310331
A;Accession: T31070

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-2531 <SHE>

A;Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 65.5%; Score 91; DB 2; Length 2531;
Best Local Similarity 53.8%; Pred. No. 2.9e-05;
Matches 14; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 6 CXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1526 CLERYGNGFCDBECNNIGCLYDGLDC 1551

RESULT 5

A49175

Notch B protein - mouse (fragment)

N;Alternate names: Notch homolog

C;Species: Mus musculus (house mouse)

C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C;Accession: A49175; PH1570; S32113
R;Lardelli, M.; Lendahl, U.

Exp. Cell Res. 204, 364-372, 1993

A;Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety

A;Reference number: A49175; MUID:93178563; PMID:8440332

A;Accession: A49175

A;Status: preliminary; nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-1203 <LAR>

A;Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990

A;Experimental source: embryo

A;Note: sequence extracted from NCBI backbone (NCBIP:126158)

C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision between

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;143-174/Domain: EGF homology <EGX1>

F;482-513/Domain: EGF homology <EGF1>

F;560-591/Domain: EGF homology <EGF>

F;674-705/Domain: EGF homology <EGX2>

F;712-743/Domain: EGF homology <EGF3>

F;836-867/Domain: EGF homology <EGX3>

Query Match 63.3%; Score 88; DB 2; Length 1203;

Best Local Similarity 45.2%; Pred. No. 4.5e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 1108 CQGYCQADKARDGICDEACNSHACQWDGDC 1138

RESULT 6

B49175
Notch A protein - mouse (fragment)
N;Alternate names: Notch homolog
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C;Accession: B49175; PH1569; S32109
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: B49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-387 <LAR>
A;Cross-references: EMBL:X68278; NID:g287987; FIDN:CAA48339.1; PID:g287988
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C;Superfamily: notch protein; ankyrin repeat controlling the decision between
F;27-58/Domain: EGF homology <EGF>
F;73-104/Domain: EGF homology <EGF2>
F;151-185/Domain: EGF homology <EGF1>

Query Match 61.2%; Score 85; DB 2; Length 387;
Best Local Similarity 45.2%; Pred. No. 5.3e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 289 CELPECQVDAGNKVCNLQCNNHACGWDGDC 319

RESULT 7

A48825
Notch homolog Notch protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
C;Accession: A48825
R;Reaume, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992
A;Title: Expression analysis of a Notch homologue in the mouse embryo.
A;Reference number: A48825; MUID:93050801; PMID:1426644
A;Accession: A48825
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-861 <REA>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:119144)
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;26-57/Domain: EGF homology <EGF>
F;64-95/Domain: EGF homology <EGX1>
F;198-229/Domain: EGF homology <EGF2>
F;441-472/Domain: EGF homology <EGX2>

Query Match 61.2%; Score 85; DB 2; Length 861;
Best Local Similarity 45.2%; Pred. No. 9.4e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 630 CELPECQVDAGNKVCNLQCNNHACGWDGDC 660

RESULT 8

notch protein homolog - rat
S18188
C;Species: Rattus norvegicus (Norway rat)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
C;Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A;Title: A homolog of Drosophila Notch expressed during mammalian development.
A;Reference number: S18188; MUID:92111383; PMID:1764995
A;Accession: S18188
A;Molecule type: mRNA
A;Residues: 1-2531 <WEI>
A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;987-1018/Domain: EGF homology <EGF1>
F;1025-1056/Domain: EGF homology <EGF>
F;1233-1264/Domain: EGF homology <EGF2>
F;1917-1949/Domain: ankyrin repeat homology <AN1>
F;1950-1982/Domain: ankyrin repeat homology <AN2>
F;1984-2016/Domain: ankyrin repeat homology <AN3>
F;2017-2049/Domain: ankyrin repeat homology <AN4>
F;2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 61.2%; Score 85; DB 2; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.0002;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 1449 CELPECQVDAGNKVCNLQCNNHACGWDGDC 1479

RESULT 9

A46019
notch-1 protein - mouse
N;Alternate names: notch protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003
C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Gri
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A;Reference number: A46019; MUID:93194170; PMID:8449489
A;Accession: A46019
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531
A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
R;Franco del Amo, F.; Smith, D.B.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A;Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugges
A;Reference number: S25144
A;Accession: S25144
A;Molecule type: mRNA
A;Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A;Cross-references: EMBL:Z11886
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: C49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1161-1547 <LAR>
A;Cross-references: EMBL:X68278; NID:g287987; FIDN:CAA48339.1; PID:g287988
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
R;Kopan, R.; Weintraub, H.
J. Cell Biol. 121, 631-641, 1993
A;Title: Mouse notch: expression in hair follicles correlates with cell fate determinat
A;Reference number: A46438; MUID:93252998; PMID:8486742

A/Accession: B46438
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1865-1932, 'RR', 1935-1937, 'L', 1938-1967, 'I', 1969-2044, 'IE', 2047-2052, 'S', 2054
A/Experimental source: embryo
A/Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBI:131247)
C/Comment: This protein has many EGF repeats and 1n-12[1172]/Notch repeats.
C/Genetics: This protein is one of the neurogenic proteins controlling the decision between
A/Gene: notch-1
A/Map position: 2
A/Note: proximal region of chromosome 2
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:106-138/Domain: EGF homology <EGF1>
F:144-175/Domain: EGF homology <EG01>
F:222-254/Domain: EGF homology <EGF2>
F:261-292/Domain: EGF homology <EG02>
F:339-370/Domain: EGF homology <EG03>
F:416-449/Domain: EGF homology <EGF3>
F:456-487/Domain: EGF homology <EG04>
F:494-525/Domain: EGF homology <EG05>
F:532-563/Domain: EGF homology <EG06>
F:607-638/Domain: EGF homology <EG07>
F:682-713/Domain: EGF homology <EG08>
F:757-788/Domain: EGF homology <EG09>
F:795-826/Domain: EGF homology <EG10>
F:873-904/Domain: EGF homology <EG11>
F:911-942/Domain: EGF homology <EG12>
F:949-980/Domain: EGF homology <EG13>
F:987-1018/Domain: EGF homology <EG14>
F:1025-1056/Domain: EGF homology <EG15>
F:1063-1094/Domain: EGF homology <EG16>
F:1149-1180/Domain: EGF homology <EG17>
F:1187-1218/Domain: EGF homology <EG18>
F:1233-1264/Domain: EGF homology <EGF4>
F:1352-1383/Domain: EGF homology <EG19>
F:1391-1425/Domain: EGF homology <EGF>
F:1917-1948/Domain: ankyrin repeat homology <AN1>
F:1949-1981/Domain: ankyrin repeat homology <AN2>
F:1983-2015/Domain: ankyrin repeat homology <AN3>
F:2016-2048/Domain: ankyrin repeat homology <AN4>
F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 61.2%; Score 85; DB 2; Length 2511;
Best Local Similarity 45.2%; Pred. No. 0.0002;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1449 CELPEQVDAGNKVCNLQCNHACGWDGDC 1479

RESULT 10
A49128
cell-fate determining gene Notch2 protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
C/Accession: A49128
R/Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A/Title: Notch2: a second mammalian Notch gene.
A/Reference number: A49128; MUID:93202015; PMID:1295745
A/Accession: A49128
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-2471 <WEI>
A/Experimental source: Schwann cell
A/Note: sequence extracted from NCBI backbone (NCBI:127811)
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:264-295/Domain: EGF homology <EGX1>
F:799-830/Domain: EGF homology <EGF1>
F:877-908/Domain: EGF homology <EGX2>
F:1029-1060/Domain: EGF homology <EGF>

F:1067-1098/Domain: EGF homology <EGX3>
F:1153-1184/Domain: EGF homology <EGF3>
F:1191-1222/Domain: EGF homology <EGX4>
F:1876-1908/Domain: ankyrin repeat homology <AN1>
F:1909-1941/Domain: ankyrin repeat homology <AN2>
F:1943-1975/Domain: ankyrin repeat homology <AN3>
F:1976-2008/Domain: ankyrin repeat homology <AN4>
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 60.4%; Score 84; DB 2; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.00028;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1425 CLSQYCADKARDGICDEACNSHACQWDGDC 1455

RESULT 11
A40043
notch protein homolog TAN-1 precursor - human
C/Species: Homo sapiens (man)
C/Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 02-Aug-2002
C/Accession: A40043
R/Elissen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar
Cell 66, 649-661, 1991
A/Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
A/Reference number: A40043; MUID:91347367; PMID:1831692
A/Accession: A40043
A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr
A/Molecule type: mRNA
A/Residues: 1-2555 <ELI>
A/Cross-references: GB:M73980
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:261-292/Domain: EGF homology <EGX1>
F:494-525/Domain: EGF homology <EGF1>
F:987-1018/Domain: EGF homology <EGX2>
F:1149-1180/Domain: EGF homology <EGF>
F:1187-1218/Domain: EGF homology <EGF3>
F:1233-1264/Domain: EGF homology <EGF3>
F:1927-1959/Domain: ankyrin repeat homology <AN1>
F:1960-1992/Domain: ankyrin repeat homology <AN2>
F:1994-2026/Domain: ankyrin repeat homology <AN3>
F:2027-2059/Domain: ankyrin repeat homology <AN4>
F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 60.4%; Score 84; DB 2; Length 2555;
Best Local Similarity 45.2%; Pred. No. 0.00028;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1449 CELPEQEDAGNKVCSLQCNHACGWDGDC 1479

RESULT 12
S78549
notch3 protein - human
C/Species: Homo sapiens (man)
C/Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002
C/Accession: S78549; S71825
R/Joutel, A.; Tournier-Lasserre, E.
submitted to the EMBL Data Library, April 1997
A/Reference number: S78549
A/Accession: S78549
A/Molecule type: mRNA
A/Residues: 1-2321 <JOU1>
A/Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592
R/Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabryat, H.; Mouton, P.; Alamowitz
X, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.
Nature 383, 707-710, 1996
A/Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke
A/Reference number: S71825; MUID:97032728; PMID:8878478

A:Accession: S71825
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 67-113;138-194;268-333, 'G', 335-346;536-613;716-765;1240-1279;1815-1888 <JOU2
A:Cross-references: EMBL:U97669
C:Genetics:
A:Gene: notch3
A:Map position: 19p13.1
C:Function:
A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: tandem repeat; transmembrane protein
F:123-155/Domain: EGF homology <EGX1>
F:162-194/Domain: EGF homology <EGF1>
F:240-271/Domain: EGF homology <EGX2>
F:318-349/Domain: EGF homology <EGF>
F:473-504/Domain: EGF homology <EGX3>
F:853-884/Domain: EGF homology <EGF3>
F:928-959/Domain: EGF homology <EGX4>
F:1838-1870/Domain: ankyrin repeat homology <AN1>
F:1871-1903/Domain: ankyrin repeat homology <AN2>
F:1905-1937/Domain: ankyrin repeat homology <AN3>
F:1938-1970/Domain: ankyrin repeat homology <AN4>
F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 59.7%; Score 83; DB 2; Length 2321;
Best Local Similarity 41.9%; Pred. No. 0.00037;
Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1387 CPRAACQAKRGDQRCDCRCSNPGCGWDGDC 1417

RESULT 13

S45306

notch 3 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002

C:Accession: S45306

Riardelli, M.; Dahlstrand, J.; Lendahl, U.

Mech. Dev. 46, 123-136, 1994

A:Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-

A:Reference number: S45306; MUID:95001556; PMID:7918097

A:Accession: S45306

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2318 <LAR>

A:Cross-references: EMBL:X74760; NID:g483580; PIDN:CA52776.1; PID:g483581

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

F:163-195/Domain: EGF homology <EGF1>

F:474-505/Domain: EGF homology <EGF>

F:854-885/Domain: EGF homology <EGF2>

F:1839-1871/Domain: ankyrin repeat homology <AN1>

F:1872-1904/Domain: ankyrin repeat homology <AN2>

F:1906-1938/Domain: ankyrin repeat homology <AN3>

F:1939-1971/Domain: ankyrin repeat homology <AN4>

F:1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 59.0%; Score 82; DB 2; Length 2318;
Best Local Similarity 41.9%; Pred. No. 0.00051;
Matches 13; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1388 CPRAACQAKRGDQRCDCRCSNPGCGWDGDC 1418

RESULT 14

A35844

Xotch protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 02-Aug-2002

C:Accession: A35844
R:Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A:Title: Xotch, the Xenopus homolog of Drosophila notch.
A:Reference number: A35844; MUID:90385285; PMID:2402639
A:Accession: A35844
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr.
A:Molecule type: mRNA
A:Residues: 1-2524 <COF>
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: transmembrane protein
F:146-177/Domain: EGF homology <EGX1>
F:184-215/Domain: EGF homology <EGF1>
F:222-254/Domain: EGF homology <EGF>
F:456-487/Domain: EGF homology <EGX2>
F:757-788/Domain: EGF homology <EGF3>
F:1025-1056/Domain: EGF homology <EGX3>
F:1924-1956/Domain: ankyrin repeat homology <AN1>
F:1957-1989/Domain: ankyrin repeat homology <AN2>
F:1991-2023/Domain: ankyrin repeat homology <AN3>
F:2024-2056/Domain: ankyrin repeat homology <AN4>
F:2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 57.6%; Score 80; DB 2; Length 2524;
Best Local Similarity 41.9%; Pred. No. 0.001;
Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
Db 1448 CENEQCESELADNKCVCNANCNNHACGWDGDC 1478

RESULT 15

A32901

gfp1 protein precursor - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Dec-1989 #sequence_revision 04-Sep-1992 #text_change 20-Sep-1999

C:Accession: A32901

R:Yochem, J.; Greenwald, I.

Cell 58, 553-563, 1989

A:Title: glp-1 and lin-12, genes implicated in distinct cell-cell interactions in Caeno-

A:Reference number: A32901; MUID:89336787; PMID:2758466

A:Accession: A32901

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1295 <YOC>

A:Cross-references: GB:M25580; NID:g156316; PIDN:AAA28058.1; PID:g156317

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C:Keywords: transmembrane protein

F:121-151/Domain: EGF homology <EGF1>

F:275-307/Domain: EGF homology <EGX1>

F:373-405/Domain: EGF homology <EGF>

F:411-442/Domain: EGF homology <EGF3>

F:450-478/Domain: EGF homology <EGF2>

F:961-993/Domain: ankyrin repeat homology <AN1>

F:1074-1106/Domain: ankyrin repeat homology <AN2>

F:1107-1139/Domain: ankyrin repeat homology <AN3>

Query Match 56.8%; Score 79; DB 2; Length 1295;
Best Local Similarity 46.2%; Pred. No. 0.00089;
Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 6 CXKXGNGXCDXXCNNAACXXDXDC 31
Db 542 CADQFANGVCNQECCNNECLYDGLDC 567

Search completed: May 19, 2004, 16:00:37
Job time : 2.29658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:51:04 ; Search time 0.785805 Seconds
(without alignments)
2054.168 Million cell updates/sec

Title: US-09-903-199-4
Perfect score: 139
Sequence: 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	75.5	2703	1	NOTC DROME
2	97	69.8	2437	1	NTC1_BRARE
3	92	66.2	1429	1	LII2_CABEL
4	88	63.3	2470	1	NTC2_MOUSE
5	85	61.2	2531	1	NTC1_MOUSE
6	85	61.2	2531	1	NTC1_MOUSE
7	84	60.4	2471	1	NTC2_HUMAN
8	84	60.4	2471	1	NTC2_HUMAN
9	84	60.4	2556	1	NTC1_MOUSE
10	83	59.7	2319	1	NTC3_HUMAN
11	83	59.7	2321	1	NTC3_HUMAN
12	82	59.0	2318	1	NTC3_MOUSE
13	80	57.6	2524	1	NOTC_MOUSE
14	79	56.8	1295	1	GLP1_CABEL
15	79	56.8	1964	1	NTC4_MOUSE
16	73	52.5	2003	1	NTC4_HUMAN
17	61	43.9	67	1	MTA_SPHGR
18	61	43.9	67	1	MTA_SPHGR
19	60	43.2	1627	1	PAPA_HUMAN
20	53	38.1	68	1	MT_LYPTI
21	50	36.0	64	1	MT_STENE
22	50	36.0	64	1	MT_STENE
23	48.5	34.9	65	1	ECAR_STRPU
24	47.5	34.2	65	1	ECAR_STRPU
25	46.5	33.5	571	1	MT_PARLI
26	46	33.1	575	1	TRBM_HUMAN
27	45	32.4	64	1	MTA_STRPU
28	45	32.4	72	1	DISI_CROAT
29	45	32.4	73	1	DISI_CROCC
30	45	32.4	73	1	DISI_CROVL
31	45	32.4	73	1	DISI_SISBA
32	45	32.4	73	1	DISI_SISCT
33	45	32.4	787	1	ITB6_MOUSE

34	45	32.4	2907	1	FBN2_MOUSE	Q61555 mus musculus
35	45	32.4	2911	1	FBN2_HUMAN	P35556 homo sapien
36	44.5	32.0	788	1	ITB6_HUMAN	P18564 homo sapien
37	44	31.7	84	1	SCX9_CENSC	Q95wc9 centruroid
38	44	31.7	186	1	AGI3_WHEAT	P10969 triticum ae
39	43.5	31.3	610	1	LEM2_HUMAN	P16581 homo sapien
40	43	30.9	316	1	CHIB_POPTR	P16061 populus tri
41	43	30.9	4753	1	LRP_CAEEL	Q04833 caenorhabdi
42	42.5	30.6	873	1	LDVR_MOUSE	P98156 mus musculus
43	42.5	30.6	873	1	LDVR_RABIT	P35953 cryptolagus
44	42.5	30.6	873	1	LDVR_RAT	P98166 rattus norv
45	42.5	30.6	2871	1	FBN1_BOVIN	P98133 bos taurus

ALIGNMENTS

RESULT 1
NOTC DROME STANDARD; PRT; 2703 AA.
AC P07207; O97458; P04154; Q9W4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurogenic locus Notch protein precursor.
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats.";
RL Cell 43:567-581 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors.";
RL Mol. Cell. Biol. 6:3094-3108 (1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fostler C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko F., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Mnana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaackle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsova A., Henderson N.S.,
RA McMillan P.J., Sales C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of *D.*
RT *melanogaster*.";
RL Science 287:2220-2222(2000).
[5]
RP SEQUENCE OF 2505-2611 FROM N.A.
RX MEDLINE=85099329; PubMed=2981631;
RA Wharton K.A., Yedvobnick B., Finnerly V.G., Artavanis-Tsakonas S.;
RT "opa: a novel family of transcribed repeats shared by the Notch locus
RT and other developmentally regulated loci in *D. melanogaster*.";
RL Cell 40:55-62(1985).
[6]
RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE=87257846; PubMed=3037327;
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
RT "Restriction of P-element insertions at the Notch locus of *Drosophila*
RT *melanogaster*.";
RL Mol. Cell. Biol. 7:1545-1548(1987).
[7]
RP INTERACTION WITH DX, AND MUTANT SU42C.
RX MEDLINE=94215489; PubMed=8162848;
RA Diederich R.J., Matsuno K., Hing H., Artavanis-Tsakonas S.;
RT "Cytosolic interaction between *delta*tex and Notch ankyrin repeats
RT implicates *delta*tex in the Notch signaling pathway.";
RL Development 120:473-481(1994).
[8]
RP INTERACTION WITH DX.
RX MEDLINE=95401878; PubMed=7671825;
RA Matsuno K., Diederich R.J., Go M.J., Blaumueller C.M.,
RA Artavanis-Tsakonas S.;
RT "Delta ϵ tex acts as a positive regulator of Notch signaling through
RT interactions with the Notch ankyrin repeats.";
RL Development 121:2633-2644(1995).
[9]
RP S3 CLEAVAGE BY PSN.
RX MEDLINE=99221487; PubMed=10206646;
RA Struhl G., Greenwald I.;
RT "Presenilin is required for activity and nuclear access of Notch in
RT *Drosophila*.";
RL Nature 398:522-525(1999).
[10]
RP S3 CLEAVAGE BY PSN.
RX MEDLINE=99221488; PubMed=10206647;
RL

```

RA Ye Y., Lukinova N., Fortini M.E.;
RT "Neurogenic phenotypes and altered Notch processing in Drosophila
RT Presentilin mutants.";
RL Nature 398:525-529(1999).
RN [11]
RP S2 CLEAVAGE BY KUZ.
RX MEDLINE=21657146; PubMed=11799064;
RA Lieber T., Kidd S., Young M.W.;
RT "kuzbanian-mediated cleavage of Drosophila Notch.";
RL Genes Dev. 16:209-221(2002).
RN [12]
RP MUTANT MCD5.
RX MEDLINE=21575956; PubMed=11719214;
RA Ramain P., Khéchumian K., Seugnet L., Arbogast N., Ackermann C.,
RA Heitzler P.;
RT "Novel Notch alleles reveal a Deltex-dependent pathway repressing
RT neural fate.";
RL Curr. Biol. 11:1729-1738(2001).
RN [13]
RP REVIEW.
RX MEDLINE=22256570; PubMed=12369105;
RA Portin P.;
RT "General outlines of the molecular genetics of the Notch signalling
RT pathway in Drosophila melanogaster: a review.";
RL Hereditas 136:89-96(2002).
CC -1- FUNCTION: Signaling protein, which regulates, with both positive
CC and negative signals, the differentiation of at least central and
CC peripheral nervous system and eye, wing disk, oogenesis, segmental
CC appendages such as antennae and legs, and muscles, through lateral
CC inhibition or induction. Functions as a receptor for membrane-
CC bound ligands Delta and Serrate to regulate cell-fate
CC determination. Upon ligand activation, and releasing from the cell
CC membrane, the Notch intracellular domain (NICD) forms a
CC transcriptional activator complex with Su(H) (Suppressor of
CC hairless) and activates genes of the E(spl) complex. Essential for
CC proper differentiation of ectoderm.
CC -1- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dx
CC via its ANK repeats.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and
CC S3 cleavage, it is released from the cell membrane and enters into
CC the nucleus in conjunction with Su(H).
CC -1- PTM: Upon binding its ligands such as Delta or Serrate, it is
CC cleaved (S2 cleavage) in its extracellular domain, close to the
CC transmembrane domain. S2 cleavage is probably mediated by Kuz. It
CC is then cleaved (S3 cleavage) downstream of its transmembrane
CC domain, releasing it from the cell membrane. S3 cleavage requires
CC Pan.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16152; AAB59220.1; -
DR EMBL; M16153; AAB59220.1; JOINED.
DR EMBL; M16149; AAB59220.1; JOINED.
DR EMBL; M16150; AAB59220.1; JOINED.
DR EMBL; M16151; AAB59220.1; JOINED.
DR EMBL; K03508; AAA28725.1; -
DR EMBL; M13689; AAA28725.1; JOINED.
DR EMBL; K03507; AAA28725.1; JOINED.
DR EMBL; AE003426; AAF45848.2; -
DR EMBL; AL035436; CAB37610.1; -
DR EMBL; AL035395; CAB37610.1; JOINED.
DR EMBL; M12175; AAA74496.1; -
DR EMBL; M16025; AAA28726.1; -

```

```

Query Match          75.5%;   Score 105;   DB 1;   Length 2703;
Best Local Similarity 54.8%;   Pred. No. 7.1e-08;
Matches 17;   Conservative 0;   Mismatches 14;   Indels 0;   Gaps 0;

Db          1482   CDKRGCTEKQNGNGICDSDCNTYACNFDGND 1512

QY          1   CDXXXXXXKKXGNXGXCDDXCNNMACXXDXDC 31
              ||| ||| ||| ||| ||| ||| ||| |||
DB          1482   CDKRGCTEKQNGNGICDSDCNTYACNFDGND 1512

RESULT 2
NTCL_BRARE
ID   NTCL_BRARE          STANDARD;          PRT;   2437 AA.
AC   P46530;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 1 precursor.
GN   NOTCH1A OR NOTCH.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=94128602; PubMed=8297791;
RA   Bierkamp C., Campos-Ortega J.A.;
RT   "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT   its pattern of transcription during early embryogenesis.";
RL   Mech. Dev. 43:87-100(1993).
CC   -!- FUNCTION: Implicated in cell fate specifications during
CC   embryo development. May be involved in the formation of the
CC   neural plate, notochord and brain vesicles.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC   stages. During gastrulation is differentially expressed,
CC   accumulating predominantly in the prechordal mesoderm and
CC   notochord. At the end of gastrulation, expressed along the
CC   anterior-posterior axis including the developing neural plate
CC   and differentiating mesoderm. Also present in the developing
CC   brain and head regions.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X69088; CAA48831.1; -.
DR   PIR; S42612; S42612.
DR   HSSP; P00740; 1EDM.
DR   ZFIN; ZDB-GENE-990415-173; notch1a.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_5.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001861; EGF_2a.
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 36.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.

```

[illegible]

FT	DISULFID	67	86	BY SIMILARITY
FT	DISULFID	88	97	BY SIMILARITY
FT	DISULFID	105	116	BY SIMILARITY
FT	DISULFID	110	126	BY SIMILARITY
FT	DISULFID	128	137	BY SIMILARITY
FT	DISULFID	143	154	BY SIMILARITY
FT	DISULFID	148	163	BY SIMILARITY
FT	DISULFID	165	174	BY SIMILARITY
FT	DISULFID	181	194	BY SIMILARITY
FT	DISULFID	188	203	BY SIMILARITY
FT	DISULFID	205	214	BY SIMILARITY
FT	DISULFID	221	232	BY SIMILARITY
FT	DISULFID	226	242	BY SIMILARITY
FT	DISULFID	244	253	BY SIMILARITY
FT	DISULFID	260	271	BY SIMILARITY
FT	DISULFID	265	280	BY SIMILARITY
FT	DISULFID	282	291	BY SIMILARITY
FT	DISULFID	298	311	BY SIMILARITY
FT	DISULFID	305	320	BY SIMILARITY
FT	DISULFID	322	331	BY SIMILARITY
FT	DISULFID	338	349	BY SIMILARITY
FT	DISULFID	343	358	BY SIMILARITY
FT	DISULFID	360	369	BY SIMILARITY
FT	DISULFID	375	386	BY SIMILARITY
FT	DISULFID	380	397	BY SIMILARITY
FT	DISULFID	399	408	BY SIMILARITY
FT	DISULFID	415	428	BY SIMILARITY
FT	DISULFID	422	437	BY SIMILARITY
FT	DISULFID	439	448	BY SIMILARITY
FT	DISULFID	455	466	BY SIMILARITY
FT	DISULFID	460	475	BY SIMILARITY
FT	DISULFID	477	486	BY SIMILARITY
FT	DISULFID	493	503	BY SIMILARITY
FT	DISULFID	498	512	BY SIMILARITY
FT	DISULFID	514	523	BY SIMILARITY
FT	DISULFID	530	541	BY SIMILARITY
FT	DISULFID	535	550	BY SIMILARITY
FT	DISULFID	552	561	BY SIMILARITY
FT	DISULFID	568	578	BY SIMILARITY
FT	DISULFID	573	587	BY SIMILARITY
FT	DISULFID	589	598	BY SIMILARITY
FT	DISULFID	605	616	BY SIMILARITY
FT	DISULFID	610	625	BY SIMILARITY
FT	DISULFID	627	636	BY SIMILARITY
FT	DISULFID	643	653	BY SIMILARITY
FT	DISULFID	648	662	BY SIMILARITY
FT	DISULFID	664	673	BY SIMILARITY
FT	DISULFID	680	691	BY SIMILARITY
FT	DISULFID	685	700	BY SIMILARITY
FT	DISULFID	702	711	BY SIMILARITY
FT	DISULFID	718	728	BY SIMILARITY
FT	DISULFID	723	737	BY SIMILARITY
FT	DISULFID	739	748	BY SIMILARITY
FT	DISULFID	755	766	BY SIMILARITY
FT	DISULFID	760	775	BY SIMILARITY
FT	DISULFID	777	786	BY SIMILARITY
FT	DISULFID	793	804	BY SIMILARITY
FT	DISULFID	798	813	BY SIMILARITY
FT	DISULFID	815	824	BY SIMILARITY
FT	DISULFID	831	842	BY SIMILARITY
FT	DISULFID	836	853	BY SIMILARITY
FT	DISULFID	855	864	BY SIMILARITY
FT	DISULFID	871	882	BY SIMILARITY
FT	DISULFID	876	891	BY SIMILARITY
FT	DISULFID	893	902	BY SIMILARITY
FT	DISULFID	909	920	BY SIMILARITY

```

Query Match          69.8%; Score 97; DB 1; Length 2437;
Best Local Similarity 48.4%; Pred. No. 9.7e-07;
Matches 15; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY      1 CDXXXCXKKXGNGXCDXXCNNAACXXDXDC 31
      |: | : | | | | | | | | | |

```

```

Db      1447  CE1A0CEGRRGNAICDTQCNNHACGWDGDC 1477

RESULT 3
L112_CAEEL
ID      L112_CAEEL      STANDARD;      PRT;      1429 AA.
AC      P14585;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Lin-12 protein precursor.
        Lin-12 OR R107.8.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RX      MEDLINE=88334747; PubMed=3419531;
RA      Yochem J., Weston K., Greenwald I.;
RA      "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
RT      protein with overall similarity to Drosophila Notch.";
RL      Nature 335:547-550 (1988).
        [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RX      MEDLINE=94150718; PubMed=7906398;
RA      Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA      Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA      Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA      Fulton L., Gardner A., Green P., Hawkes T., Hillier L., Jier M.,
RA      Johnston L., Jones M., Kershaw P., Kirsten J., Laister N.,
RA      Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA      Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA      Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA      Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA      Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA      Woldman P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans.";
RL      Nature 368:32-38 (1994).
CC      -1- FUNCTION: Involved in several cell fates decisions that requires
CC      cell-cell interactions. It is possible that lin-12 encodes a
CC      membrane-bound receptor for a signal that enables expression of
CC      the ventral uterine precursor cell fate.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
CC      -1- SIMILARITY: Contains 13 EGF-like domains.
CC      -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC      -1- SIMILARITY: Contains 5 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M12069; AAA70191.1; -.
DR      EMBL; Z14092; CAA78474.1; -.
DR      PIR; S06434; S06434.
DR      HSSP; P00740; 1EDM.
DR      WormPep; R107.8; CE00274.
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR      InterPro; IPR000742; EGF_2.
DR      InterPro; IPR001881; EGF_Ca.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00023; ank; 6.

```


RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -|- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC -|- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=035516-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=035516-2; Sequence=VSP_001405;
 CC Note=No experimental confirmation available;
 CC -|- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 CC neuroepithelia, somites, optic vesicles and branchial arches, but
 CC not heart.
 CC -|- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 CC the postnatal ependymal cells, and the choroid plexus throughout
 CC embryonic and postnatal development.
 CC -|- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -|- PTM: Phosphorylated.
 CC -|- SIMILARITY: Belongs to the NOTCH family.
 CC -|- SIMILARITY: Contains 35 EGF-like domains.
 CC -|- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -|- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D32210; BAA22094.1; -;
 DR EMBL; X68279; CAA48340.1; -;
 DR EMBL; U31881; AAC52924.1; -;
 DR PIR; A49175; A49175.

DR HSP; P16109; 1FSB.
 DR MGD; MGI:97364; Notch2.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR00152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBL00.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5.
 FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
 FT DOMAIN 258 294 EGF-LIKE 7.
 FT DOMAIN 296 334 EGF-LIKE 8.
 FT DOMAIN 336 372 EGF-LIKE 9.
 FT DOMAIN 373 411 EGF-LIKE 10.
 FT DOMAIN 413 452 EGF-LIKE 11.
 FT DOMAIN 454 490 EGF-LIKE 12.
 FT DOMAIN 492 528 EGF-LIKE 13.
 FT DOMAIN 530 566 EGF-LIKE 14.
 FT DOMAIN 568 603 EGF-LIKE 15.
 FT DOMAIN 605 641 EGF-LIKE 16.
 FT DOMAIN 643 678 EGF-LIKE 17.
 FT DOMAIN 680 716 EGF-LIKE 18.
 FT DOMAIN 718 753 EGF-LIKE 19.
 FT DOMAIN 755 791 EGF-LIKE 20.
 FT DOMAIN 793 829 EGF-LIKE 21.
 FT DOMAIN 831 869 EGF-LIKE 22.
 FT DOMAIN 871 907 EGF-LIKE 23.
 FT DOMAIN 909 945 EGF-LIKE 24.
 FT DOMAIN 947 983 EGF-LIKE 25.
 FT DOMAIN 985 1021 EGF-LIKE 26.
 FT DOMAIN 1023 1059 EGF-LIKE 27.
 FT DOMAIN 1061 1097 EGF-LIKE 28.
 FT DOMAIN 1099 1145 EGF-LIKE 29.
 FT DOMAIN 1147 1183 EGF-LIKE 30.
 FT DOMAIN CALCIUM-BINDING (POTENTIAL).

```

FT DOMAIN 1185 1221 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1223 1260 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1262 1300 EGF-LIKE 33.
FT DOMAIN 1302 1345 EGF-LIKE 34.
FT DOMAIN 1372 1410 EGF-LIKE 35.
FT REPEAT 1418 1454 LIN/NOTCH 1.
FT REPEAT 1501 1533 LIN/NOTCH 2.
FT REPEAT 1825 1869 ANK 1.

```

```

Query Match 63.3%; Score 88; DB 1; Length 2470;
Best Local Similarity 45.2%; Pred. No. 2.1e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

```

```

OY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
DB 1423 COSQYCADKARDGICDEACNSHACQWDGDC 1453

```

RESULT 5

```

NTCL MOUSE STANDARD; PRT; 2531 AA.
ID NTCL MOUSE
AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)
DE (MT14) (p300).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.";
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE=93050801; PubMed=1426644;
RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo.";
RL Dev. Biol. 154:377-387(1992).
RN [3]
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.";
RL Development 115:737-744(1992).
RN [4]
RP SEQUENCE OF 1161-1547 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [5]
RP SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE=99364499; PubMed=10437788;
RA Lee J.S., Ishimoto A., Yanagawa S.I.;
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RL FEBS Lett. 455:276-280(1999).

```

```

RN [6]
RP SEQUENCE OF 1950-2201 FROM N.A.
RX MEDLINE=98029496; PubMed=9384671;
RA Messerle M., Folio M., Nehls M., Eggert H., Boehm T.;
RT "Dynamic changes in gene expression during in vitro differentiation of
RT mouse embryonic stem cells.";
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN [7]
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FUFIN-LIKE CONVERTASE, AND
RP MUTAGENESIS OF 1651-ARG--ARG-1654.
RX MEDLINE=98318619; PubMed=9653148;
RA Logeat F., Bessia C., Brou C., LeBail O., Jarrault S., Seidah N.G.,
RA Israel A.;
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RT convertase.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN [8]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [9]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN [10]
RP INTERACTION WITH DTX1 AND DTX2.
RX MEDLINE=21123790; PubMed=11226752;
RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Aravanis-Tsakonas S.,
RA Okano H., Matsuno K.;
RT "Murine homologs of delta define a novel gene family involved in
RT vertebrate Notch signaling and neurogenesis.";
RL Int. J. Dev. Neurosci. 19:21-35(2001).
RN [11]
RP FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somite formation and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds. Interacts with DTX1 and DTX2.
CC SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytic processing NICD is translocated to the nucleus.
CC ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q01705-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC Note=No experimental confirmation available;
CC TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.
CC DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC the neuroepithelium. At 13.5 dpc expressed in the surface
CC ectoderm, eye and developing whisker follicles.
CC PTM: Synthesized in the endoplasmic reticulum as an inactive form

```

CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z11886; CAA77941.1; -.
 DR EMBL; L02613; AAK14898.1; -.
 DR EMBL; X68278; CAA48339.1; -.
 DR EMBL; AJ238029; CAB40733.1; -.
 DR EMBL; X82562; CAA57909.1; -.
 DR PIR; A46019; A46019.
 DR PIR; B49175; B49175.
 DR HSSP; P00740; 1EDM.
 DR MGD; MGI:97363; Notch1.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0030154; P:cell differentiation; IMP.
 DR GO; GO:0007386; P:compartment specification; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 7.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT CHAIN 1744 2531

FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
 Query Match 61.2%; Score 85; DB 1; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 5.9e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXKXKXGNGXCDXXCNNAACXDXGDC 31
 Db 1449 CELPEQVDAGNKVCNLCNNHACGWDGDC 1479
 RESULT 6
 NTCL_RAT
 ID NTCL_RAT STANDARD; PRT; 2531 AA.
 AC Q07008;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Schwann cell;
 RX MEDLINE=92111383; PubMed=1764995;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RT development.";
 RL Development 113:199-205(1991).
 RN [2]
 RP REVISIONS TO 1652-1653.
 RA Weinmaster G.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55(2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By

```

CC similarity).
CC -1- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC Expressed in postnatal central nervous system (CNS) germinal zones
CC and, in early postnatal life, within numerous cells throughout the
CC CNS. Found in both subventricular and ventricular germinal zones.
CC -1- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
CC days 12 and 14 and decrease rapidly to much lower levels in the
CC adult.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57405; CAA40667.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 25.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 18
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION (BY
FT CHAIN 1744 2531 SIMILARITY).
FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN (BY
FT DOMAIN 19 1723 SIMILARITY).
FT TRANSMEM 1724 1746 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1747 2531 POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).

```

FT	DOMAIN	20	58	EGF-LIKE 1.
FT	DOMAIN	59	99	EGF-LIKE 2.
FT	DOMAIN	102	139	EGF-LIKE 3.
FT	DOMAIN	140	176	EGF-LIKE 4.
FT	DOMAIN	178	216	EGF-LIKE 5.
FT	DOMAIN	218	255	EGF-LIKE 6.
FT	DOMAIN	257	293	EGF-LIKE 7.
FT	DOMAIN	295	333	EGF-LIKE 8.
FT	DOMAIN	335	371	EGF-LIKE 9.
FT	DOMAIN	372	410	EGF-LIKE 10.
FT	DOMAIN	412	450	EGF-LIKE 11.
FT	DOMAIN	452	488	EGF-LIKE 12.
FT	DOMAIN	490	526	EGF-LIKE 13.
FT	DOMAIN	528	564	EGF-LIKE 14.
FT	DOMAIN	566	601	EGF-LIKE 15.
FT	DOMAIN	603	639	EGF-LIKE 16.
FT	DOMAIN	641	676	EGF-LIKE 17.
FT	DOMAIN	678	714	EGF-LIKE 18.
FT	DOMAIN	716	751	EGF-LIKE 19.
FT	DOMAIN	753	789	EGF-LIKE 20.
FT	DOMAIN	791	827	EGF-LIKE 21.
FT	DOMAIN	829	867	EGF-LIKE 22.
FT	DOMAIN	869	905	EGF-LIKE 23.
FT	DOMAIN	907	943	EGF-LIKE 24.
FT	DOMAIN	945	981	EGF-LIKE 25.
FT	DOMAIN	983	1019	EGF-LIKE 26.
FT	DOMAIN	1021	1057	EGF-LIKE 27.
FT	DOMAIN	1059	1095	EGF-LIKE 28.
FT	DOMAIN	1097	1143	EGF-LIKE 29.
FT	DOMAIN	1145	1181	EGF-LIKE 30.
FT	DOMAIN	1183	1219	EGF-LIKE 31.
FT	DOMAIN	1221	1265	EGF-LIKE 32.
FT	DOMAIN	1267	1305	EGF-LIKE 33.
FT	DOMAIN	1307	1346	EGF-LIKE 34.
FT	DOMAIN	1348	1384	EGF-LIKE 35.
FT	DOMAIN	1387	1426	EGF-LIKE 36.
FT	REPEAT	1445	1480	LIN/NOTCH 1.
FT	REPEAT	1481	1522	LIN/NOTCH 2.
FT	REPEAT	1523	1562	LIN/NOTCH 3.
FT	REPEAT	1917	1946	ANK 1.
FT	REPEAT	1950	1980	ANK 2.
FT	REPEAT	1984	2013	ANK 3.
FT	REPEAT	2017	2046	ANK 4.
FT	REPEAT	2050	2079	ANK 5.
FT	DOMAIN	1730	1733	POLY-ALA.
FT	DOMAIN	1891	1894	POLY-GLU.
FT	DOMAIN	2258	2261	POLY-PRO.
FT	DOMAIN	2497	2500	POLY-SER.
FT	SITE	1654	1655	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY
FT	DISULFID	24	37	SIMILARITY).
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.

Query Match 61.2%; Score 85; DB 1; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 5.9e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGXCDXXCNNAACXXDXDC 31

Db 1449 CELPEQEDAGNKVCNLQCNHACGWDGDC 1479

RESULT 7

NTC2_HUMAN STANDARD; PRT; 2471 AA.

ID NTC2_HUMAN 004721; Q9H240;

AC 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).

GN NOTCH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Blaumüller C.M., Mann R.S.;

RT "Complete human notch 2 (hN2) cDNA sequence.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast tumor;

RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.G.;

RT "Human Notch2, a novel member of cell-fate determining NOTCH family.";

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 967-1229 FROM N.A.

RC TISSUE=T-cell;

RA Lemasson I., Devaux C., Mesnard J.M.;

RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 1810-2447 FROM N.A.

RC TISSUE=Brain;

RA MEDLINE=93265135; PubMed=1303260;

RA Stifani S., Blaumüller C.M., Redhead N.J., Hill R.E., Artavanis-Tsakonas S.;

RT "Human homologs of a Drosophila enhancer of split gene product define a novel family of nuclear proteins.";

RL Nat. Genet. 2:119-127(1992).

RN [5]

RP POST-TRANSLATIONAL PROCESSING.

RA MEDLINE=97386453; PubMed=9244302;

RA Blaumüller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;

RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.";

RL Cell 90:281-291(1997).

RN [6]

RP IDENTIFICATION OF LIGANDS.

RA MEDLINE=99180765; PubMed=10079256;

RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L., Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

RT "Human ligands of the Notch receptor.";

RL Am. J. Pathol. 154:785-794(1999).

CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta1 to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity).

CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.

CC -1- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung, skeletal muscle and liver.

CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

CC -1- PTM: Phosphorylated (By similarity).

CC -1- SIMILARITY: Belongs to the NOTCH family.

CC -1- SIMILARITY: Contains 35 EGF-like domains.

CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.

CC -1- SIMILARITY: Contains 6 ANK repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF308601; AAA36377.2; -.

DR EMBL; AF315356; AAG37073.1; -.

DR EMBL; U77493; AAB19224.1; -.

DR HSSP; P00740; 1EDM.

DR Genew; HGNC:7882; NOTCH2.

DR MIM; 600275; -.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF_II.

DR InterPro; IPR006209; EGF_III.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR008297; Notch.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00008; EGF; 35.

DR Pfam; PF00066; notch; 2.

DR PIRSF; PIRSF002279; Notch; 1.

DR PRINTS; PR00010; EGFBL00D.

DR PRINTS; PR00011; EGF_LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SMO0248; ANK; 6.

DR SMART; SMO0179; EGF_CA; 23.

DR SMART; SMO0004; NL; 2.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS50088; ANK_REPEAT; 4.

DR PROSITE; PS00010; ASX_HYDROXYL; 22.

DR PROSITE; PS00022; EGF_1; 34.

DR PROSITE; PS01186; EGF_2; 29.

DR PROSITE; PS50026; EGF_3; 35.

DR PROSITE; PS01187; EGF_CA; 22.

DR Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

FT SIGNAL 1 25

FT CHAIN 26 2471

FT CHAIN 1666 2471

FT CHAIN 1697 2471

FT DOMAIN 26 1677

FT TRANSMEM 1678 1698

FT DOMAIN 1699 2471

FT DOMAIN 26 63

FT DOMAIN 64 102

FT DOMAIN 105 143

FT DOMAIN 144 180

FT EGF-LIKE 1.

FT EGF-LIKE 2.

FT EGF-LIKE 3.

FT EGF-LIKE 4.

FT NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).

FT NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).

FT CYTOPLASMIC (POTENTIAL).

FT EGF-LIKE 1.

FT EGF-LIKE 2.

FT EGF-LIKE 3.

FT EGF-LIKE 4.


```

FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 221 258 EGF-LIKE 6.
FT DOMAIN 260 296 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 298 336 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 338 374 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 375 413 EGF-LIKE 10.
FT DOMAIN 415 454 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 456 492 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 494 530 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 532 568 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 570 605 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 607 643 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 645 680 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 682 718 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 720 755 EGF-LIKE 19.
FT DOMAIN 757 793 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 795 831 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 833 871 EGF-LIKE 22.
FT DOMAIN 873 909 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 911 947 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 949 985 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 987 1023 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1025 1061 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1063 1099 EGF-LIKE 28.
FT DOMAIN 1101 1147 EGF-LIKE 29.
FT DOMAIN 1149 1185 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1187 1223 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1225 1262 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1264 1302 EGF-LIKE 33.
FT DOMAIN 1304 1343 EGF-LIKE 34.
FT DOMAIN 1374 1412 EGF-LIKE 35.
FT REPEAT 1420 1456 LIN/NOTCH 1.
FT REPEAT 1503 1535 LIN/NOTCH 2.
FT REPEAT 1827 1871 ANK 1.
FT REPEAT 1876 1905 ANK 2.
FT REPEAT 1909 1939 ANK 3.
FT REPEAT 1943 1972 ANK 4.
FT REPEAT 1976 2005 ANK 5.
FT REPEAT 2009 2038 ANK 6.
FT DOMAIN 1645 1648 POLY-ALA.
FT DOMAIN 1994 1997 POLY-LEU.
FT DOMAIN 2426 2429 POLY-SER.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 90 BY SIMILARITY.
FT DISULFID 92 101 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 115 131 BY SIMILARITY.
FT DISULFID 133 142 BY SIMILARITY.
FT DISULFID 148 159 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 170 179 BY SIMILARITY.
FT DISULFID 186 198 BY SIMILARITY.
FT DISULFID 192 207 BY SIMILARITY.
FT DISULFID 209 218 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 248 257 BY SIMILARITY.

```

Query Match 60.4%; Score 84; DB 1; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 8.1e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Qy 1 CDXXXGXXKXGNGXCDXXCNNAACXXDXDC 31
 Db 1425 CLSGYCADKARDGVCEACNSHACQWDGDC 1455

RESULT 8
 NTC2_RAT
 ID_NTC2_RAT STANDARD; PRT; 2471 AA.

```

AC Q9QW30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
GN NOTCH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941 (1992).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irwin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain development.";
RT J. Comp. Neurol. 436:167-181 (2001).
RL
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs. May play an essential role in postimplantation development, probably in some aspect of cell specification and/or differentiation (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.
CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL, M93661; AAK13558.1; -.
CC PIR, A49128; A49128.

```

```

DR HSSP; P00743; 1CCF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 2.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS50026; EGF_3; 35.
DR PROSITE; PS01187; EGF_CA; 22.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT CHAIN 1697 2471
FT TRANSMEM 1678 1697
FT TRANSMEM 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 258
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343

```

```

FT DOMAIN 1374 1412 EGF-LIKE 35.
FT DOMAIN 1645 1648 POLY-ALA.
FT DOMAIN 1994 1997 POLY-LEU.
FT DOMAIN 2426 2429 POLY-SER.
FT DOMAIN 2446 2451 POLY-GLY.
FT REPEAT 1420 1456 LIN/NOTCH 1.
FT REPEAT 1503 1535 LIN/NOTCH 2.
FT REPEAT 1827 1871 ANK 1.
FT REPEAT 1876 1905 ANK 2.
FT REPEAT 1909 1939 ANK 3.
FT REPEAT 1943 1972 ANK 4.
FT REPEAT 1976 2005 ANK 5.
FT REPEAT 2009 2038 ANK 6.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 90 BY SIMILARITY.
FT DISULFID 92 101 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 115 131 BY SIMILARITY.
FT DISULFID 133 142 BY SIMILARITY.
FT DISULFID 148 159 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 170 179 BY SIMILARITY.
FT DISULFID 186 198 BY SIMILARITY.
FT DISULFID 192 207 BY SIMILARITY.
FT DISULFID 209 218 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 248 257 BY SIMILARITY.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.

```

Query Match 60.4%; Score 84; DB 1; Length 2471;
 Best local Similarity 45.2%; Pred. No. 8.1e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXKXXKXGNGXCDXXCNNAACXXDXDC 31
 DB 1425 CLSQYCADKARDGICDEACNSHACQWDGDC 1455

RESULT 9
 NTCL HUMAN STANDARD; PRT; 2556 AA.
 AC P46531;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hnl)
 DE (Translocation-associated notch protein TAN1).
 GN NOTCH1 OR TAN1.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 CC NCBT_TaxID=9606;
 CC [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Mann R.S., Blaumüller C.M., Zagouras P.;
 RT "Complete human notch 1 (hN1) cDNA sequence."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE OF 1-2444 FROM N.A.
 RX MEDLINE=91347367; PubMed=1831692;
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RA Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms."
 RL Cell 66:649-661(1991).
 RN [3]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor."
 RL Am. J. Pathol. 154:785-794(1999).
 RN [4]
 RP INTERACTION WITH DTX1.
 RX MEDLINE=98250176; PubMed=9590294;
 RA Matsuno K., Eastman D., Mitsiadis T., Quinn A.M., Carcangiu M.L.,
 RA Ordentlich P., Kadesch T., Artavanis-Tsakonas S.;
 RT "Human deltex is a conserved regulator of Notch signalling."
 RL Nat. Genet. 19:74-78(1998).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Tagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May be important for normal lymphocyte
 CC function. In altered form, may contribute to transformation or
 CC progression in some T-cell neoplasms. Involved in the maturation
 CC of both CD4+ and CD8+ cells in the thymus (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity). Interacts with DTX1 and DTX2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
 CC brain stem and lung. Also present in most adult tissues where it
 CC is found mainly in lymphoid tissues.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- DISEASE: NOTCH1 truncation is associated with T-cell acute
 CC lymphoblastic leukemia.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF308602; AAC33848.1; -.
 DR EMBL; M73980; AAA60614.1; -.
 DR HSSP; P00740; 1EDM.
 DR Genew; HGNC:7881; NOTCH1.
 DR MIM; 190198; -.
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0003793; F:defense/immunity protein activity; NAS.
 DR GO; GO:0006955; P:immune response; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_II.
 DR InterPro; IPR002049; laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 20.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1..18
 FT CHAIN 19..256
 FT CHAIN 1722..2556
 FT CHAIN 1755..2556
 FT DOMAIN 19..1736
 FT TRANSMEM 1737..1757
 FT DOMAIN 1758..2556
 FT DOMAIN 20..58
 FT DOMAIN 59..99
 FT DOMAIN 102..139
 FT DOMAIN 140..176
 FT DOMAIN 178..216
 FT DOMAIN 218..255
 FT DOMAIN 257..293
 FT DOMAIN 295..333
 FT DOMAIN 335..371
 FT DOMAIN 372..410
 FT DOMAIN 412..450
 FT DOMAIN 452..488
 FT DOMAIN 490..526
 FT DOMAIN 528..564
 FT DOMAIN 566..601
 FT DOMAIN 603..639
 FT DOMAIN 641..676
 FT DOMAIN 678..714
 FT DOMAIN 716..751
 FT DOMAIN 753..789
 FT DOMAIN 791..827
 FT DOMAIN 829..868
 FT DOMAIN 868..922
 FT DOMAIN 924..956
 FT DOMAIN 958..1000
 FT DOMAIN 1002..1040
 FT DOMAIN 1042..1080
 FT DOMAIN 1082..1120
 FT DOMAIN 1122..1160
 FT DOMAIN 1162..1200
 FT DOMAIN 1202..1240
 FT DOMAIN 1242..1280
 FT DOMAIN 1282..1320
 FT DOMAIN 1322..1360
 FT DOMAIN 1362..1400
 FT DOMAIN 1402..1440
 FT DOMAIN 1442..1480
 FT DOMAIN 1482..1520
 FT DOMAIN 1522..1560
 FT DOMAIN 1562..1600
 FT DOMAIN 1602..1640
 FT DOMAIN 1642..1680
 FT DOMAIN 1682..1720
 FT DOMAIN 1722..1760
 FT DOMAIN 1762..1800
 FT DOMAIN 1802..1840
 FT DOMAIN 1842..1880
 FT DOMAIN 1882..1920
 FT DOMAIN 1922..1960
 FT DOMAIN 1962..2000
 FT DOMAIN 2002..2040
 FT DOMAIN 2042..2080
 FT DOMAIN 2082..2120
 FT DOMAIN 2122..2160
 FT DOMAIN 2162..2200
 FT DOMAIN 2202..2240
 FT DOMAIN 2242..2280
 FT DOMAIN 2282..2320
 FT DOMAIN 2322..2360
 FT DOMAIN 2362..2400
 FT DOMAIN 2402..2440
 FT DOMAIN 2442..2480
 FT DOMAIN 2482..2520
 FT DOMAIN 2522..2560
 FT DOMAIN 2562..2600
 FT DOMAIN 2602..2640
 FT DOMAIN 2642..2680
 FT DOMAIN 2682..2720
 FT DOMAIN 2722..2760
 FT DOMAIN 2762..2800
 FT DOMAIN 2802..2840
 FT DOMAIN 2842..2880
 FT DOMAIN 2882..2920
 FT DOMAIN 2922..2960
 FT DOMAIN 2962..3000
 FT DOMAIN 3002..3040
 FT DOMAIN 3042..3080
 FT DOMAIN 3082..3120
 FT DOMAIN 3122..3160
 FT DOMAIN 3162..3200
 FT DOMAIN 3202..3240
 FT DOMAIN 3242..3280
 FT DOMAIN 3282..3320
 FT DOMAIN 3322..3360
 FT DOMAIN 3362..3400
 FT DOMAIN 3402..3440
 FT DOMAIN 3442..3480
 FT DOMAIN 3482..3520
 FT DOMAIN 3522..3560
 FT DOMAIN 3562..3600
 FT DOMAIN 3602..3640
 FT DOMAIN 3642..3680
 FT DOMAIN 3682..3720
 FT DOMAIN 3722..3760
 FT DOMAIN 3762..3800
 FT DOMAIN 3802..3840
 FT DOMAIN 3842..3880
 FT DOMAIN 3882..3920
 FT DOMAIN 3922..3960
 FT DOMAIN 3962..4000
 FT DOMAIN 4002..4040
 FT DOMAIN 4042..4080
 FT DOMAIN 4082..4120
 FT DOMAIN 4122..4160
 FT DOMAIN 4162..4200
 FT DOMAIN 4202..4240
 FT DOMAIN 4242..4280
 FT DOMAIN 4282..4320
 FT DOMAIN 4322..4360
 FT DOMAIN 4362..4400
 FT DOMAIN 4402..4440
 FT DOMAIN 4442..4480
 FT DOMAIN 4482..4520
 FT DOMAIN 4522..4560
 FT DOMAIN 4562..4600
 FT DOMAIN 4602..4640
 FT DOMAIN 4642..4680
 FT DOMAIN 4682..4720
 FT DOMAIN 4722..4760
 FT DOMAIN 4762..4800
 FT DOMAIN 4802..4840
 FT DOMAIN 4842..4880
 FT DOMAIN 4882..4920
 FT DOMAIN 4922..4960
 FT DOMAIN 4962..5000
 FT DOMAIN 5002..5040
 FT DOMAIN 5042..5080
 FT DOMAIN 5082..5120
 FT DOMAIN 5122..5160
 FT DOMAIN 5162..5200
 FT DOMAIN 5202..5240
 FT DOMAIN 5242..5280
 FT DOMAIN 5282..5320
 FT DOMAIN 5322..5360
 FT DOMAIN 5362..5400
 FT DOMAIN 5402..5440
 FT DOMAIN 5442..5480
 FT DOMAIN 5482..5520
 FT DOMAIN 5522..5560
 FT DOMAIN 5562..5600
 FT DOMAIN 5602..5640
 FT DOMAIN 5642..5680
 FT DOMAIN 5682..5720
 FT DOMAIN 5722..5760
 FT DOMAIN 5762..5800
 FT DOMAIN 5802..5840
 FT DOMAIN 5842..5880
 FT DOMAIN 5882..5920
 FT DOMAIN 5922..5960
 FT DOMAIN 5962..6000
 FT DOMAIN 6002..6040
 FT DOMAIN 6042..6080
 FT DOMAIN 6082..6120
 FT DOMAIN 6122..6160
 FT DOMAIN 6162..6200
 FT DOMAIN 6202..6240
 FT DOMAIN 6242..6280
 FT DOMAIN 6282..6320
 FT DOMAIN 6322..6360
 FT DOMAIN 6362..6400
 FT DOMAIN 6402..6440
 FT DOMAIN 6442..6480
 FT DOMAIN 6482..6520
 FT DOMAIN 6522..6560
 FT DOMAIN 6562..6600
 FT DOMAIN 6602..6640
 FT DOMAIN 6642..6680
 FT DOMAIN 6682..6720
 FT DOMAIN 6722..6760
 FT DOMAIN 6762..6800
 FT DOMAIN 6802..6840
 FT DOMAIN 6842..6880
 FT DOMAIN 6882..6920
 FT DOMAIN 6922..6960
 FT DOMAIN 6962..7000
 FT DOMAIN 7002..7040
 FT DOMAIN 7042..7080
 FT DOMAIN 7082..7120
 FT DOMAIN 7122..7160
 FT DOMAIN 7162..7200
 FT DOMAIN 7202..7240
 FT DOMAIN 7242..7280
 FT DOMAIN 7282..7320
 FT DOMAIN 7322..7360
 FT DOMAIN 7362..7400
 FT DOMAIN 7402..7440
 FT DOMAIN 7442..7480
 FT DOMAIN 7482..7520
 FT DOMAIN 7522..7560
 FT DOMAIN 7562..7600
 FT DOMAIN 7602..7640
 FT DOMAIN 7642..7680
 FT DOMAIN 7682..7720
 FT DOMAIN 7722..7760
 FT DOMAIN 7762..7800
 FT DOMAIN 7802..7840
 FT DOMAIN 7842..7880
 FT DOMAIN 7882..7920
 FT DOMAIN 7922..7960
 FT DOMAIN 7962..8000
 FT DOMAIN 8002..8040
 FT DOMAIN 8042..8080
 FT DOMAIN 8082..8120
 FT DOMAIN 8122..8160
 FT DOMAIN 8162..8200
 FT DOMAIN 8202..8240
 FT DOMAIN 8242..8280
 FT DOMAIN 8282..8320
 FT DOMAIN 8322..8360
 FT DOMAIN 8362..8400
 FT DOMAIN 8402..8440
 FT DOMAIN 8442..8480
 FT DOMAIN 8482..8520
 FT DOMAIN 8522..8560
 FT DOMAIN 8562..8600
 FT DOMAIN 8602..8640
 FT DOMAIN 8642..8680
 FT DOMAIN 8682..8720
 FT DOMAIN 8722..8760
 FT DOMAIN 8762..8800
 FT DOMAIN 8802..8840
 FT DOMAIN 8842..8880
 FT DOMAIN 8882..8920
 FT DOMAIN 8922..8960
 FT DOMAIN 8962..9000
 FT DOMAIN 9002..9040
 FT DOMAIN 9042..9080
 FT DOMAIN 9082..9120
 FT DOMAIN 9122..9160
 FT DOMAIN 9162..9200
 FT DOMAIN 9202..9240
 FT DOMAIN 9242..9280
 FT DOMAIN 9282..9320
 FT DOMAIN 9322..9360
 FT DOMAIN 9362..9400
 FT DOMAIN 9402..9440
 FT DOMAIN 9442..9480
 FT DOMAIN 9482..9520
 FT DOMAIN 9522..9560
 FT DOMAIN 9562..9600
 FT DOMAIN 9602..9640
 FT DOMAIN 9642..9680
 FT DOMAIN 9682..9720
 FT DOMAIN 9722..9760
 FT DOMAIN 9762..9800
 FT DOMAIN 9802..9840
 FT DOMAIN 9842..9880
 FT DOMAIN 9882..9920
 FT DOMAIN 9922..9960
 FT DOMAIN 9962..10000
 FT DOMAIN 10002..10040
 FT DOMAIN 10042..10080
 FT DOMAIN 10082..10120
 FT DOMAIN 10122..10160
 FT DOMAIN 10162..10200
 FT DOMAIN 10202..10240
 FT DOMAIN 10242..10280
 FT DOMAIN 10282..10320
 FT DOMAIN 10322..10360
 FT DOMAIN 10362..10400
 FT DOMAIN 10402..10440
 FT DOMAIN 10442..10480
 FT DOMAIN 10482..10520
 FT DOMAIN 10522..10560
 FT DOMAIN 10562..10600
 FT DOMAIN 10602..10640
 FT DOMAIN 10642..10680
 FT DOMAIN 10682..10720
 FT DOMAIN 10722..10760
 FT DOMAIN 10762..10800
 FT DOMAIN 10802..10840
 FT DOMAIN 10842..10880
 FT DOMAIN 10882..10920
 FT DOMAIN 10922..10960
 FT DOMAIN 10962..11000
 FT DOMAIN 11002..11040
 FT DOMAIN 11042..11080
 FT DOMAIN 11082..11120
 FT DOMAIN 11122..11160
 FT DOMAIN 11162..11200
 FT DOMAIN 11202..11240
 FT DOMAIN 11242..11280
 FT DOMAIN 11282..11320
 FT DOMAIN 11322..11360
 FT DOMAIN 11362..11400
 FT DOMAIN 11402..11440
 FT DOMAIN 11442..11480
 FT DOMAIN 11482..11520
 FT DOMAIN 11522..11560
 FT DOMAIN 11562..11600
 FT DOMAIN 11602..11640
 FT DOMAIN 11642..11680
 FT DOMAIN 11682..11720
 FT DOMAIN 11722..11760
 FT DOMAIN 11762..11800
 FT DOMAIN 11802..11840
 FT DOMAIN 11842..11880
 FT DOMAIN 11882..11920
 FT DOMAIN 11922..11960
 FT DOMAIN 11962..12000
 FT DOMAIN 12002..12040
 FT DOMAIN 12042..12080
 FT DOMAIN 12082..12120
 FT DOMAIN 12122..12160
 FT DOMAIN 12162..12200
 FT DOMAIN 12202..12240
 FT DOMAIN 12242..12280
 FT DOMAIN 12282..12320
 FT DOMAIN 12322..12360
 FT DOMAIN 12362..12400
 FT DOMAIN 12402..12440
 FT DOMAIN 12442..12480
 FT DOMAIN 12482..12520
 FT DOMAIN 12522..12560
 FT DOMAIN 12562..12600
 FT DOMAIN 12602..12640
 FT DOMAIN 12642..12680
 FT DOMAIN 12682..12720
 FT DOMAIN 12722..12760
 FT DOMAIN 12762..12800
 FT DOMAIN 12802..12840
 FT DOMAIN 12842..12880
 FT DOMAIN 12882..12920
 FT DOMAIN 12922..12960
 FT DOMAIN 12962..13000
 FT DOMAIN 13002..13040
 FT DOMAIN 13042..13080
 FT DOMAIN 13082..13120
 FT DOMAIN 13122..13160
 FT DOMAIN 13162..13200
 FT DOMAIN 13202..13240
 FT DOMAIN 13242..13280
 FT DOMAIN 13282..13320
 FT DOMAIN 13322..13360
 FT DOMAIN 13362..13400
 FT DOMAIN 13402..13440
 FT DOMAIN 13442..13480
 FT DOMAIN 13482..13520
 FT DOMAIN 13522..13560
 FT DOMAIN 13562..13600
 FT DOMAIN 13602..13640
 FT DOMAIN 13642..13680
 FT DOMAIN 13682..13720
 FT DOMAIN 13722..13760
 FT DOMAIN 13762..13800
 FT DOMAIN 13802..13840
 FT DOMAIN 13842..13880
 FT DOMAIN 13882..13920
 FT DOMAIN 13922..13960
 FT DOMAIN 13962..14000
 FT DOMAIN 14002..14040
 FT DOMAIN 14042..14080
 FT DOMAIN 14082..14120
 FT DOMAIN 14122..14160
 FT DOMAIN 14162..14200
 FT DOMAIN 14202..14240
 FT DOMAIN 14242..14280
 FT DOMAIN 14282..14320
 FT DOMAIN 14322..14360
 FT DOMAIN 14362..14400
 FT DOMAIN 14402..14440
 FT DOMAIN 14442..14480
 FT DOMAIN 14482..14520
 FT DOMAIN 14522..14560
 FT DOMAIN 14562..14600
 FT DOMAIN 14602..14640
 FT DOMAIN 14642..14680
 FT DOMAIN 14682..14720
 FT DOMAIN 14722..14760
 FT DOMAIN 14762..14800
 FT DOMAIN 14802..14840
 FT DOMAIN 14842..14880
 FT DOMAIN 14882..14920
 FT DOMAIN 14922..14960
 FT DOMAIN 14962..15000
 FT DOMAIN 15002..15040
 FT DOMAIN 15042..15080
 FT DOMAIN 15082..15120
 FT DOMAIN 15122..15160
 FT DOMAIN 15162..15200
 FT DOMAIN 15202..15240
 FT DOMAIN 15242..15280
 FT DOMAIN 15282..15320
 FT DOMAIN 15322..15360
 FT DOMAIN 15362..15400
 FT DOMAIN 15402..15440
 FT DOMAIN 15442..15480
 FT DOMAIN 15482..15520
 FT DOMAIN 15522..15560
 FT DOMAIN 15562..15600
 FT DOMAIN 15602..15640
 FT DOMAIN 15642..15680
 FT DOMAIN 15682..15720
 FT DOMAIN 15722..15760
 FT DOMAIN 15762..15800
 FT DOMAIN 15802..15840
 FT DOMAIN 15842..15880
 FT DOMAIN 15882..15920
 FT DOMAIN 15922..15960
 FT DOMAIN 15962..16000
 FT DOMAIN 16002..16040
 FT DOMAIN 16042..16080
 FT DOMAIN 16082..16120
 FT DOMAIN 16122..16160
 FT DOMAIN 16162..16200
 FT DOMAIN 16202..16240
 FT DOMAIN 16242..16280
 FT DOMAIN 16282..16320
 FT DOMAIN 16322..16360
 FT DOMAIN 16362..16400
 FT DOMAIN 16402..16440
 FT DOMAIN 16442..16480
 FT DOMAIN 16482..16520
 FT DOMAIN 16522..16560
 FT DOMAIN 16562..16600
 FT DOMAIN 16602..16640
 FT DOMAIN 16642..16680
 FT DOMAIN 16682..16720
 FT DOMAIN 16722..16760
 FT DOMAIN 16762..16800
 FT DOMAIN 16802..16840
 FT DOMAIN 16842..16880
 FT DOMAIN 16882..16920
 FT DOMAIN 16922..16960
 FT DOMAIN 16962..17000
 FT DOMAIN 17002..17040
 FT DOMAIN 17042..17080
 FT DOMAIN 17082..17120
 FT DOMAIN 17122..17160
 FT DOMAIN 17162..17200
 FT DOMAIN 17202..17240
 FT DOMAIN 17242..17280
 FT DOMAIN 17282..17320
 FT DOMAIN 17322..17360
 FT DOMAIN 17362..17400
 FT DOMAIN 17402..17440
 FT DOMAIN 17442..17480
 FT DOMAIN 17482..17520
 FT DOMAIN 17522..17560
 FT DOMAIN 17562..17600
 FT DOMAIN 17602..17640
 FT DOMAIN 17642..17680
 FT DOMAIN 17682..17720
 FT DOMAIN 17722..17760
 FT DOMAIN 17762..17800
 FT DOMAIN 17802..17840
 FT DOMAIN 17842..17880
 FT DOMAIN 17882..17920
 FT DOMAIN 17922..17960
 FT DOMAIN 17962..18000
 FT DOMAIN 18002..18040
 FT DOMAIN 18042..18080
 FT DOMAIN 18082..18120
 FT DOMAIN 18122..18160
 FT DOMAIN 18162..18200
 FT DOMAIN 18202..18240
 FT DOMAIN 18242..18280
 FT DOMAIN 18282..18320
 FT DOMAIN 18322..18360
 FT DOMAIN 18362..18400
 FT DOMAIN 18402..18440
 FT DOMAIN 18442..18480
 FT DOMAIN 18482..18520
 FT DOMAIN 18522..18560
 FT DOMAIN 18562..18600
 FT DOMAIN 18602..18640
 FT DOMAIN 18642..18680
 FT DOMAIN 18682..18720
 FT DOMAIN 18722..18760
 FT DOMAIN 18762..18800
 FT DOMAIN 18802..18840
 FT DOMAIN 18842..18880
 FT DOMAIN 18882..18920
 FT DOMAIN 18922..18960
 FT DOMAIN 18962..19000
 FT DOMAIN 19002..19040
 FT DOMAIN 19042..19080
 FT DOMAIN 19082..19120
 FT DOMAIN 19122..19160
 FT DOMAIN 19162..19200
 FT DOMAIN 19202..19240
 FT DOMAIN 19242..19280
 FT DOMAIN 19282..19320
 FT DOMAIN 19322..19360
 FT DOMAIN 19362..19400
 FT DOMAIN 19402..19440
 FT DOMAIN 19442..19480
 FT DOMAIN 19482..19520
 FT DOMAIN 19522..19560
 FT DOMAIN 19562..19600
 FT DOMAIN 19602..19640
 FT DOMAIN 19642..19680
 FT DOMAIN 19682..19720
 FT DOMAIN 19722..19760
 FT DOMAIN 19762..19800
 FT DOMAIN 19802..19840
 FT DOMAIN 19842..19880
 FT DOMAIN 19882..19920
 FT DOMAIN 19922..19960
 FT DOMAIN 19962..20000
 FT DOMAIN 20002..20040
 FT DOMAIN 20042..20080
 FT DOMAIN 20082..20120
 FT DOMAIN 20122..20160
 FT DOMAIN 20162..20200
 FT DOMAIN 20202..20240
 FT DOMAIN 20242..20280
 FT DOMAIN 20282..20320
 FT DOMAIN 20322..20360
 FT DOMAIN 20362..20400
 FT DOMAIN 20402..20440
 FT DOMAIN 20442..20480
 FT DOMAIN 20482..20520
 FT DOMAIN 20522..20560
 FT DOMAIN 20562..20600
 FT DOMAIN 20602..20640
 FT DOMAIN 20642..20680
 FT DOMAIN 20682..20720
 FT DOMAIN 20722..20760
 FT DOMAIN 20762..20800
 FT DOMAIN 20802..20840
 FT DOMAIN 20842..20880
 FT DOMAIN 20882..20920
 FT DOMAIN 20922..20960
 FT DOMAIN 20962..21000
 FT DOMAIN 21002..21040
 FT DOMAIN 21042..21080
 FT DOMAIN 21082..21120
 FT DOMAIN 21122..21160
 FT DOMAIN 21162..21200
 FT DOMAIN 21202..21240
 FT DOMAIN 21242..21280
 FT DOMAIN 21282..21320
 FT DOMAIN 21322..21360
 FT DOMAIN 21362..21400
 FT DOMAIN 21402..21440
 FT DOMAIN 21442..21480
 FT DOMAIN 21482..21520
 FT DOMAIN 21522..21560
 FT DOMAIN 21562..21600
 FT DOMAIN 21602..21640
 FT DOMAIN 21642..21680
 FT DOMAIN 21682..21720
 FT DOMAIN 21722..21760
 FT DOMAIN 21762..21800
 FT DOMAIN 21802..21840
 FT DOMAIN 21842..21880
 FT DOMAIN 21882..21920
 FT DOMAIN 21922..21960
 FT DOMAIN 21962..22000
 FT DOMAIN 22002..22040
 FT DOMAIN 22042..22080
 FT DOMAIN 22082..22120
 FT DOMAIN 22122..22160
 FT DOMAIN 22162..22200
 FT DOMAIN 22202..22240
 FT DOMAIN 22242..22280
 FT DOMAIN 22282..22320
 FT DOMAIN 22322..22360
 FT DOMAIN 22362..22400
 FT DOMAIN 22402..22440
 FT DOMAIN 22442..22480
 FT DOMAIN 22482..22520
 FT DOMAIN 22522..22560
 FT DOMAIN 22562..22600
 FT DOMAIN 22602..22640
 FT DOMAIN 22642..22680
 FT DOMAIN 22682..22720
 FT DOMAIN 22722..22760
 FT DOMAIN 22762..22800
 FT DOMAIN 22802..22840
 FT DOMAIN

```

FT DOMAIN 870 906 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL) .
FT DOMAIN 908 944 EGF-LIKE 24.
FT DOMAIN 946 982 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL) .
FT DOMAIN 984 1020 EGF-LIKE 26.
FT DOMAIN 1022 1058 EGF-LIKE 27.
FT DOMAIN 1060 1096 EGF-LIKE 28.
FT DOMAIN 1098 1144 EGF-LIKE 29.
FT DOMAIN 1146 1182 EGF-LIKE 30.
FT DOMAIN 1184 1220 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL) .
FT DOMAIN 1222 1266 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL) .
FT DOMAIN 1268 1306 EGF-LIKE 33.
FT DOMAIN 1308 1347 EGF-LIKE 34.
FT DOMAIN 1349 1385 EGF-LIKE 35.
FT DOMAIN 1388 1427 EGF-LIKE 36.
FT REPEAT 1446 1481 LIN/NOTCH 1.
FT REPEAT 1482 1523 LIN/NOTCH 2.
FT REPEAT 1524 1563 LIN/NOTCH 3.
FT REPEAT 1928 1957 ANK 1.
FT REPEAT 1961 1991 ANK 2.
FT REPEAT 1995 2024 ANK 3.
FT REPEAT 2028 2057 ANK 4.
FT REPEAT 2061 2090 ANK 5.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DOMAIN 2522 2525 POLY-SER.
FT SITE 1665 1666 CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY) .
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.

Query Match 60.4%; Score 84; DB 1; Length 2556;
Best Local Similarity 45.2%; Pred. No. 8.3e-05;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31
Db 1450 CELPEQEDAGNKVCSLQCNHACGWDGDC 1480

RESULT 10
NTC3_RAT STANDARD; PRT; 2319 AA.
AC Q9RI72;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3) .
GN NOTCH3.
OS Rattus norvegicus (Rat) .
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
RT "Rattus norvegicus mRNA for Notch 3."

```

```

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate.";
RL Neuron 29:45-55 (2001).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain
RT development.";
RL J. Comp. Neurol. 436:167-181 (2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 34 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF164486; AAD46653.2; -.
CC HSSP; P00740; 1EDM.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000152; Asx_hydroxyl_S.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001438; EGF_IT.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR008297; Notch.
CC InterPro; IPR008600; Notch_dom.
CC Pfam; PF00023; ank; 6.

```

[illegible]

FT	DISULFID	50	67	BY SIMILARITY.
FT	DISULFID	69	78	BY SIMILARITY.
FT	DISULFID	84	95	BY SIMILARITY.
FT	DISULFID	89	108	BY SIMILARITY.
FT	DISULFID	110	119	BY SIMILARITY.
FT	DISULFID	125	136	BY SIMILARITY.
FT	DISULFID	130	146	BY SIMILARITY.
FT	DISULFID	148	157	BY SIMILARITY.
FT	DISULFID	164	176	BY SIMILARITY.
FT	DISULFID	170	185	BY SIMILARITY.
FT	DISULFID	187	196	BY SIMILARITY.
FT	DISULFID	203	214	BY SIMILARITY.
FT	DISULFID	208	224	BY SIMILARITY.
FT	DISULFID	226	235	BY SIMILARITY.
FT	DISULFID	242	253	BY SIMILARITY.
FT	DISULFID	247	262	BY SIMILARITY.
FT	DISULFID	264	273	BY SIMILARITY.
FT	DISULFID	280	293	BY SIMILARITY.
FT	DISULFID	287	302	BY SIMILARITY.
FT	DISULFID	304	313	BY SIMILARITY.
FT	DISULFID	320	331	BY SIMILARITY.
FT	DISULFID	325	340	BY SIMILARITY.
FT	DISULFID	342	351	BY SIMILARITY.
FT	DISULFID	357	368	BY SIMILARITY.
FT	DISULFID	362	379	BY SIMILARITY.
FT	DISULFID	381	390	BY SIMILARITY.
FT	DISULFID	397	410	BY SIMILARITY.
FT	DISULFID	404	419	BY SIMILARITY.
FT	DISULFID	421	430	BY SIMILARITY.
FT	DISULFID	437	448	BY SIMILARITY.
FT	DISULFID	442	457	BY SIMILARITY.
FT	DISULFID	459	468	BY SIMILARITY.
FT	DISULFID	475	486	BY SIMILARITY.
FT	DISULFID	480	495	BY SIMILARITY.
FT	DISULFID	497	506	BY SIMILARITY.
FT	DISULFID	513	524	BY SIMILARITY.
FT	DISULFID	518	533	BY SIMILARITY.
FT	DISULFID	535	544	BY SIMILARITY.
FT	DISULFID	551	561	BY SIMILARITY.

Query Match 59.7%; Score 83; DB 1; Length 2319;
Best Local Similarity 41.9%; Pred. No. 0.00011;
Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```
QY      1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
      | | | | | | | | | | | | | | | |
Db      1389 CPRAACQAKRGDQNCNDCRECNSPGCGWDGDC 1419
```

```

RESULT 11
NTC3_HUMAN
ID NTC3_HUMAN STANDARD; PRT; 2321 AA.
AC Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97032728; PubMed=8878487;
RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.;
RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
RT causing stroke and dementia.";
RL Nature 383:707-710(1996).
[2]

```


RP SEQUENCE FROM N.A.
RA Gunel M., Artavanis-Tsakonas S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA Tranckheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA Carrano A.V.;
RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
RT 19p13.1";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX MEDLINE=98049753; PubMed=9388399;
RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA Vayssiere C., Cruaud C., Maciazek J., Weissbach J., Bousser M.-G.,
RA Bach J.-F., Tournier-Lasserre E.;
RT "Strong clustering and stereotyped nature of Notch3 mutations in
RT CADASIL patients.";
RL Lancet 350:1511-1515 (1997).
RN [5]
RP VARIANT CADASIL 114-GLY--PRO-120 DEL.
RX MEDLINE=20264473; PubMed=10802807;
RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
RA Ruchoux M.M., Lucas C., Lays D., Bousser M.-G., Tournier-Lasserre E.;
RT "Splice site mutation causing a seven amino acid Notch3 in-frame
RT deletion in CADASIL.";
RL Neurology 54:1874-1875 (2000).
RN [6]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Leitman J., Ward D., Ieh-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794 (1999).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC jagged1, jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC tissues.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
CC dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type

CC of stroke and dementia of which key features include recurrent
CC subcortical ischemic events and vascular dementia.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 34 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U97669; AAB91371.1; --
DR EMBL; AF058900; AAC14346.1; --
DR EMBL; AF058881; AAC14346.1; JOINED.
DR EMBL; AF058882; AAC14346.1; JOINED.
DR EMBL; AF058883; AAC14346.1; JOINED.
DR EMBL; AF058884; AAC14346.1; JOINED.
DR EMBL; AF058885; AAC14346.1; JOINED.
DR EMBL; AF058886; AAC14346.1; JOINED.
DR EMBL; AF058887; AAC14346.1; JOINED.
DR EMBL; AF058888; AAC14346.1; JOINED.
DR EMBL; AF058889; AAC14346.1; JOINED.
DR EMBL; AF058890; AAC14346.1; JOINED.
DR EMBL; AF058891; AAC14346.1; JOINED.
DR EMBL; AF058892; AAC14346.1; JOINED.
DR EMBL; AF058893; AAC14346.1; JOINED.
DR EMBL; AF058894; AAC14346.1; JOINED.
DR EMBL; AF058895; AAC14346.1; JOINED.
DR EMBL; AF058896; AAC14346.1; JOINED.
DR EMBL; AF058897; AAC14346.1; JOINED.
DR EMBL; AF058898; AAC14346.1; JOINED.
DR EMBL; AF058899; AAC14346.1; JOINED.
DR EMBL; AC004257; AAC04897.1; --
DR EMBL; AC004663; AAC15789.1; ALT_INIT.
DR PIR; S78549; S78549.
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:7883; NOTCH3.
DR MIM; 600276; --
DR MIM; 125310; --
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_IT.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Lamitin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 19.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS50026; EGF_3; 34.
DR PROSITE; PS01187; EGF_CA; 16.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW Disease mutation.

FT	SIGNAL	1	39	POTENTIAL.
FT	CHAIN	40	2321	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT	CHAIN	1629	2321	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
FT	CHAIN	1662	2321	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
FT	DOMAIN	40	1643	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1644	1664	POTENTIAL.
FT	DOMAIN	1665	2321	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	40	77	EGF-LIKE 1.
FT	DOMAIN	78	118	EGF-LIKE 2.
FT	DOMAIN	119	156	EGF-LIKE 3.
FT	DOMAIN	158	195	EGF-LIKE 4.
FT	DOMAIN	197	234	EGF-LIKE 5.
FT	DOMAIN	236	272	EGF-LIKE 6.
FT	DOMAIN	274	312	EGF-LIKE 7.
FT	DOMAIN	314	350	EGF-LIKE 8.
FT	DOMAIN	351	389	EGF-LIKE 9.
FT	DOMAIN	391	429	EGF-LIKE 10.
FT	DOMAIN	431	467	EGF-LIKE 11.
FT	DOMAIN	469	505	EGF-LIKE 12.
FT	DOMAIN	507	543	EGF-LIKE 13.
FT	DOMAIN	545	580	EGF-LIKE 14.
FT	DOMAIN	582	618	EGF-LIKE 15.
FT	DOMAIN	620	655	EGF-LIKE 16.
FT	DOMAIN	657	693	EGF-LIKE 17.
FT	DOMAIN	695	730	EGF-LIKE 18.
FT	DOMAIN	734	770	EGF-LIKE 19.
FT	DOMAIN	771	808	EGF-LIKE 20.
FT	DOMAIN	810	847	EGF-LIKE 21.
FT	DOMAIN	849	885	EGF-LIKE 22.
FT	DOMAIN	887	922	EGF-LIKE 23.

```

Query Match          59.7%; Score 83; DB 1; Length 2321;
Best Local Similarity 41.9%; Pred. No. 0.00011;
Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 CDXXXXXXKXGXCDXXCNNAACXXDXDC 31
Db 1387 CPRAACQAKRGDQRCDCRCNSPGCGWDGDC 1417

RESULT 12
NTC3_MOUSE
ID NTC3_MOUSE STANDARD; PRT; 2318 AA.
AC 061982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ICR X Swiss Webster;
RX MEDLINE=95001556; PubMed=7918097;
RA Lardelli M., Dalstrand J., Lendahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium.";
RL Mech. Dev. 46:123-136(1994).
RN [2]
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.
RP MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Wntine notch homologs (NL-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [3]
RN POST-TRANSLATIONAL PROCESSING.

```

```

RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
  among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
  Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
  Upon ligand activation through the released notch intracellular
  domain (NICD) it forms a transcriptional activator complex with
  RBP-J kappa and activates genes of the enhancer of split locus.
  Affects the implementation of differentiation, proliferation and
  apoptotic programs (By similarity). May play a role during CNS
  development.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
  terminal fragment N(EC) which are probably linked by disulfide
  bonds.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
  proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Proliferating neuroepithelium.
CC -!- DEVELOPMENTAL STAGE: CNS development.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
  which is proteolytically cleaved by a furin-like convertase in the
  trans-Golgi network before it reaches the plasma membrane to yield
  an active, ligand-accessible form. Cleavage results in a C-
  terminal fragment N(TM) and a N-terminal fragment N(EC). Following
  ligand binding, it is cleaved by TNF-alpha converting enzyme
  (TACE) to yield a membrane-associated intermediate fragment called
  notch extracellular truncation (NEXT). This fragment is then
  cleaved by presenilin dependent gamma-secretase to release a
  notch-derived peptide containing the intracellular domain (NICD)
  from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 34 EGF-like domains.
CC -!- SIMILARITY: Contains 3 lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74760; CAA52776.1; -.
DR PIR; S45306; S45306.
DR HSSP; P00740; 1EDM.
DR MGD; MGI:99460; Notch3.
DR GO; GO:0005887; C:Integral to plasma membrane; IC.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007219; P:N signaling pathway; IC.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_5.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 33.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 19.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.

```

```

DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS50026; EGF_3; 34.
DR PROSITE; PS01187; EGF_CA; 16.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 39
FT CHAIN 40 2318
FT CHAIN 1630 2318
FT CHAIN 1663 2318
FT CHAIN 40 1643
FT TRANSMEM 1644 1664
FT DOMAIN 1665 2318
FT DOMAIN 2242 2261
FT DOMAIN 40 78
FT DOMAIN 79 119
FT DOMAIN 120 157
FT DOMAIN 159 196
FT DOMAIN 198 235
FT DOMAIN 237 273
FT DOMAIN 275 313
FT DOMAIN 315 351
FT DOMAIN 352 390
FT DOMAIN 392 430
FT DOMAIN 432 468
FT DOMAIN 470 506
FT DOMAIN 508 544
FT DOMAIN 546 581
FT DOMAIN 583 619
FT DOMAIN 621 656
FT DOMAIN 658 694
FT DOMAIN 696 731
FT DOMAIN 735 771
FT DOMAIN 772 809
FT DOMAIN 811 848
FT DOMAIN 850 886
FT DOMAIN 888 923
FT DOMAIN 925 961
FT DOMAIN 963 999
FT DOMAIN 1001 1035
FT DOMAIN 1037 1083
FT DOMAIN 1085 1121
FT DOMAIN 1123 1159
FT DOMAIN 1161 1204
FT DOMAIN 1206 1245
FT DOMAIN 1247 1288
FT DOMAIN 1290 1326
FT DOMAIN 1336 1374
FT DOMAIN 1388 1428
FT REPEAT 1429 1467
FT REPEAT 1468 1503
FT REPEAT 1839 1868
FT REPEAT 1872 1902
FT REPEAT 1906 1935
FT REPEAT 1939 1968
FT REPEAT 1972 2001
FT SITE 1572 1573
FT DISULFID 43 55
FT DISULFID 49 66
FT DISULFID 68 77
FT DISULFID 83 94
FT DISULFID 88 107
FT DISULFID 109 118
FT DISULFID 124 135
FT DISULFID 129 145
FT DISULFID 147 156
FT DISULFID 163 175
FT DISULFID 169 184
FT DISULFID 186 195
FT DISULFID 202 213
PROSITE; PS00010; ASX_HYDROXYL; 18.
PROSITE; PS00022; EGF_1; 33.
PROSITE; PS01186; EGF_2; 27.
PROSITE; PS50026; EGF_3; 34.
PROSITE; PS01187; EGF_CA; 16.
Receptor; Transcription regulation; Activator; Differentiation;
Developmental protein; Repeat; ANK repeat; EGF-like domain;
Transmembrane; Glycoprotein; Signal; Phosphorylation.
SIGNAL 1 39
CHAIN 40 2318
CHAIN 1630 2318
CHAIN 1663 2318
CHAIN 40 1643
TRANSMEM 1644 1664
DOMAIN 1665 2318
DOMAIN 2242 2261
DOMAIN 40 78
DOMAIN 79 119
DOMAIN 120 157
DOMAIN 159 196
DOMAIN 198 235
DOMAIN 237 273
DOMAIN 275 313
DOMAIN 315 351
DOMAIN 352 390
DOMAIN 392 430
DOMAIN 432 468
DOMAIN 470 506
DOMAIN 508 544
DOMAIN 546 581
DOMAIN 583 619
DOMAIN 621 656
DOMAIN 658 694
DOMAIN 696 731
DOMAIN 735 771
DOMAIN 772 809
DOMAIN 811 848
DOMAIN 850 886
DOMAIN 888 923
DOMAIN 925 961
DOMAIN 963 999
DOMAIN 1001 1035
DOMAIN 1037 1083
DOMAIN 1085 1121
DOMAIN 1123 1159
DOMAIN 1161 1204
DOMAIN 1206 1245
DOMAIN 1247 1288
DOMAIN 1290 1326
DOMAIN 1336 1374
DOMAIN 1388 1428
REPEAT 1429 1467
REPEAT 1468 1503
REPEAT 1839 1868
REPEAT 1872 1902
REPEAT 1906 1935
REPEAT 1939 1968
REPEAT 1972 2001
SITE 1572 1573
DISULFID 43 55
DISULFID 49 66
DISULFID 68 77
DISULFID 83 94
DISULFID 88 107
DISULFID 109 118
DISULFID 124 135
DISULFID 129 145
DISULFID 147 156
DISULFID 163 175
DISULFID 169 184
DISULFID 186 195
DISULFID 202 213

```

```

FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 225 234 BY SIMILARITY.
FT DISULFID 241 252 BY SIMILARITY.
FT DISULFID 246 261 BY SIMILARITY.
FT DISULFID 263 272 BY SIMILARITY.
FT DISULFID 279 292 BY SIMILARITY.
FT DISULFID 286 301 BY SIMILARITY.
FT DISULFID 303 312 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 324 339 BY SIMILARITY.
FT DISULFID 341 350 BY SIMILARITY.
FT DISULFID 356 367 BY SIMILARITY.
FT DISULFID 361 378 BY SIMILARITY.
FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.

Query Match 59.0%; Score 82; DB 1; Length 2318;
Best Local Similarity 41.9%; Pred. No. 0.00015;
Matches 13; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 1 CDXXXGXXGNGXCDXXCNNAACXXDXDC 31
Db 1388 CPRAACQAKRGDQNCNDCRENTPGCGWGDGC 1418

RESULT 13
NOTC_XENLA
ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC
 DR EMBL; M33874; AAB02039.1; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_IIke.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PRO0010; EGFLLOOD.
 DR PRINTS; PRO0011; EGFLAMININ.
 DR PRINTS; PRO1452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1
 FT CHAIN 19
 FT DOMAIN 20 2524
 FT TRANSMEM 20 1728
 FT DOMAIN 1729 1750
 FT DOMAIN 1751 2524
 FT DOMAIN 20 57
 FT DOMAIN 58 99
 FT DOMAIN 102 140
 FT DOMAIN 141 177
 FT DOMAIN 179 215
 FT DOMAIN 217 254
 FT DOMAIN 256 292
 FT DOMAIN 294 332
 FT DOMAIN 334 370
 FT DOMAIN 371 409
 FT DOMAIN 411 449
 FT DOMAIN 451 487
 FT DOMAIN 489 525
 FT DOMAIN 527 563
 FT DOMAIN 565 600
 FT DOMAIN 602 638
 FT DOMAIN 640 675
 FT DOMAIN 677 713
 FT DOMAIN 715 750
 FT DOMAIN 752 788
 FT DOMAIN 790 826
 FT DOMAIN 828 866
 FT DOMAIN 868 904
 FT DOMAIN 906 942
 FT DOMAIN 944 980
 FT DOMAIN 982 1018
 FT DOMAIN 1020 1056
 FT DOMAIN 1058 1094
 FT DOMAIN 1096 1142
 FT DOMAIN 1144 1180
 FT DOMAIN 1182 1218
 FT DOMAIN 1220 1264
 FT DOMAIN 1266 1304
 FT DOMAIN 1306 1346
 FT DOMAIN 1347 1383
 FT DOMAIN 1386 1424
 FT REPEAT 1441 1478
 POTENTIAL.
 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5.
 EGF-LIKE 6.
 EGF-LIKE 7.
 EGF-LIKE 8.
 EGF-LIKE 9.
 EGF-LIKE 10.
 EGF-LIKE 11.
 EGF-LIKE 12.
 EGF-LIKE 13.
 EGF-LIKE 14.
 EGF-LIKE 15.
 EGF-LIKE 16.
 EGF-LIKE 17.
 EGF-LIKE 18.
 EGF-LIKE 19.
 EGF-LIKE 20.
 EGF-LIKE 21.
 EGF-LIKE 22.
 EGF-LIKE 23.
 EGF-LIKE 24.
 EGF-LIKE 25.
 EGF-LIKE 26.
 EGF-LIKE 27.
 EGF-LIKE 28.
 EGF-LIKE 29.
 EGF-LIKE 30.
 EGF-LIKE 31.
 EGF-LIKE 32.
 EGF-LIKE 33.
 EGF-LIKE 34.
 EGF-LIKE 35.
 EGF-LIKE 36.
 LIN/NOTCH 1.

FT REPEAT 1479 1520
 FT REPEAT 1521 1560
 FT REPEAT 1876 1919
 FT REPEAT 1924 1953
 FT REPEAT 1957 1987
 FT REPEAT 1991 2020
 FT REPEAT 2024 2053
 FT REPEAT 2057 2086
 FT DISULFID 22 35
 FT DISULFID 29 45
 FT DISULFID 47 56
 FT DISULFID 62 74
 FT DISULFID 68 87
 FT DISULFID 89 98
 FT DISULFID 106 117
 FT DISULFID 111 128
 FT DISULFID 130 139
 FT DISULFID 145 156
 FT DISULFID 150 165
 FT DISULFID 167 176
 FT DISULFID 183 194
 FT DISULFID 188 203
 FT DISULFID 205 214
 FT DISULFID 221 232
 FT DISULFID 226 242
 FT DISULFID 244 253
 FT DISULFID 260 271
 FT DISULFID 265 280
 FT DISULFID 282 291
 FT DISULFID 298 311
 FT DISULFID 305 320
 FT DISULFID 322 331
 FT DISULFID 338 349
 FT DISULFID 343 358
 FT DISULFID 360 369
 FT DISULFID 375 386
 FT DISULFID 380 397
 FT DISULFID 399 408
 FT DISULFID 415 428
 FT DISULFID 422 437
 FT DISULFID 439 448
 FT DISULFID 455 466
 FT DISULFID 460 475
 FT DISULFID 477 486
 FT DISULFID 493 504
 FT DISULFID 498 513
 FT DISULFID 515 524
 FT DISULFID 531 542
 FT DISULFID 536 551
 FT DISULFID 553 562
 FT DISULFID 569 579
 FT DISULFID 574 588
 FT DISULFID 590 599
 FT DISULFID 606 617
 FT DISULFID 611 626
 FT DISULFID 628 637
 FT DISULFID 644 654
 FT DISULFID 649 663
 FT DISULFID 665 674
 FT DISULFID 681 692
 FT DISULFID 686 701
 FT DISULFID 703 712
 FT DISULFID 719 729
 FT DISULFID 724 738
 FT DISULFID 740 749
 FT DISULFID 756 767
 FT DISULFID 761 776
 FT DISULFID 778 787
 FT DISULFID 794 805
 FT DISULFID 799 814
 FT DISULFID 816 825
 FT DISULFID 832 843
 FT DISULFID 837 854
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 ANK 1.
 ANK 2.
 ANK 3.
 ANK 4.
 ANK 5.
 ANK 6.

```
FT DISULFID 856 865 BY SIMILARITY.
FT DISULFID 872 883 BY SIMILARITY.
FT DISULFID 877 892 BY SIMILARITY.
FT DISULFID 894 903 BY SIMILARITY.
FT DISULFID 910 921 BY SIMILARITY.
FT DISULFID 915 930 BY SIMILARITY.
FT DISULFID 932 941 BY SIMILARITY.
FT DISULFID 986 997 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1008 1017 BY SIMILARITY.
FT DISULFID 1024 1035 BY SIMILARITY.
FT DISULFID 1029 1044 BY SIMILARITY.
FT DISULFID 1046 1055 BY SIMILARITY.
FT DISULFID 1062 1073 BY SIMILARITY.
FT DISULFID 1067 1082 BY SIMILARITY.
FT DISULFID 1084 1093 BY SIMILARITY.
FT DISULFID 1100 1121 BY SIMILARITY.
FT DISULFID 1115 1130 BY SIMILARITY.
```

Query Match 57.6%; Score 80; DB 1; Length 2524;
Best Local Similarity 41.9%; Pred. No. 0.00032;
Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```
OY 1 CDXXXCXXKXGXCDXXCNNAACXXDXDC 31
Db 1447 CENEQCEIADNKVCNANCNNHACGWDGDC 1477
```

RESULT 14

GLP1_CABEL STANDARD; PRT; 1295 AA.

```
AC P13508;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE GLP-1 protein precursor.
GN GLP-1 OR EMB-33 OR F02A9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=89336787; PubMed=2758466;
RA Yochem J., Greenwald I.;
RT "glp-1 and lin-12, genes implicated in distinct cell-cell
RL interactions in C. elegans, encode similar transmembrane proteins.";
Cell 58:553-563(1989).
```

[2]

```
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ahnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RN Nature 368:32-38(1994).
```

[3]

```
RP DELETION OF 1174-1295.
RX MEDLINE=91351288; PubMed=1881436;
RA Mango S.E., Maine E.M., Kimble J.;
RT "Carboxy-terminal truncation activates glp-1 protein to specify
RT vulval fates in Caenorhabditis elegans.";
```

```
RL Nature 352:811-815(1991).
RN [4]
RP CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.
RX MEDLINE=93354444; PubMed=8350921;
RA Roehl H., Kimble J.;
RT "Control of cell fate in C. elegans by a GLP-1 peptide consisting
RT primarily of ankyrin repeats.";
RL Nature 364:632-635(1993).
RN [5]
RP FUNCTION.
RX MEDLINE=94208066; PubMed=8156602;
RA Mello C.C., Draper B.W., Priess J.R.;
RT "The maternal genes apx-1 and glp-1 and establishment of
RT dorsal-ventral polarity in the early C. elegans embryo.";
RL Cell 77:95-106(1994).
CC -!- FUNCTION: Involved in the specification of the cell fates of the
CC blastomeres, ABA and APA. Proper signaling by glp-1 induces ABA
CC descendants to produce anterior pharyngeal cells, and ABA
CC descendants to adopt a different fate. Contributes to the
CC establishment the dorsal-ventral axis in early embryos.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Acts on ABP development during 4-cell and
CC 12-cell stages, and on ABA development during 12-cell and 28-cell
CC stages.
CC -!- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.
CC -!- SIMILARITY: Contains 10 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M25580; AAA28058.1; -.
DR EMBL; Z19555; CAA79620.1; -.
DR EMBL; Z29116; CAA79620.1; JOINED.
DR EMBL; Z29116; CAA82373.1; -.
DR EMBL; Z19555; CAA82373.1; JOINED.
DR PIR; A32901; A32901.
DR HSSP; P00740; LEDM.
DR GerMOnline; 193793; -.
DR WormPep; F02A9.6; CE00237.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; 10.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS50026; EGF_3; 10.
DR PROSITE; PS01187; EGF_CA; 1.
DR Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
KM Glycoprotein; Signal.
FT SIGNAL 1 15
FT CHAIN 16 1295
FT DOMAIN 16 764
FT TRANSMEM 765 786
FT DOMAIN 787 1295
FT POTENTIAL.
FT GLP-1 PROTEIN.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
```

FT DOMAIN 19 58 EGF-LIKE 1.
 FT DOMAIN 117 152 EGF-LIKE 2.
 FT DOMAIN 154 190 EGF-LIKE 3.
 FT DOMAIN 190 230 EGF-LIKE 4.
 FT DOMAIN 232 269 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 271 308 EGF-LIKE 6.
 FT DOMAIN 316 359 EGF-LIKE 7.
 FT DOMAIN 369 406 EGF-LIKE 8.
 FT DOMAIN 407 443 EGF-LIKE 9.
 FT DOMAIN 446 479 EGF-LIKE 10.
 FT REPEAT 493 527 LIN/NOTCH 1.
 FT REPEAT 528 568 LIN/NOTCH 2.
 FT REPEAT 569 608 LIN/NOTCH 3.
 FT REPEAT 961 990 ANK 1.
 FT REPEAT 994 1023 ANK 2.
 FT REPEAT 1030 1062 ANK 3.
 FT REPEAT 1074 1103 ANK 4.
 FT REPEAT 1107 1136 ANK 5.
 FT DISULFID 23 35 BY SIMILARITY.
 FT DISULFID 29 46 BY SIMILARITY.
 FT DISULFID 48 57 BY SIMILARITY.
 FT DISULFID 121 131 BY SIMILARITY.
 FT DISULFID 126 140 BY SIMILARITY.
 FT DISULFID 142 151 BY SIMILARITY.
 FT DISULFID 158 169 BY SIMILARITY.
 FT DISULFID 163 178 BY SIMILARITY.
 FT DISULFID 180 189 BY SIMILARITY.
 FT DISULFID 194 206 BY SIMILARITY.
 FT DISULFID 201 218 BY SIMILARITY.
 FT DISULFID 220 229 BY SIMILARITY.
 FT DISULFID 236 248 BY SIMILARITY.
 FT DISULFID 242 257 BY SIMILARITY.
 FT DISULFID 259 268 BY SIMILARITY.
 FT DISULFID 275 286 BY SIMILARITY.
 FT DISULFID 280 296 BY SIMILARITY.
 FT DISULFID 298 307 BY SIMILARITY.
 FT DISULFID 373 384 BY SIMILARITY.
 FT DISULFID 378 394 BY SIMILARITY.
 FT DISULFID 396 405 BY SIMILARITY.
 FT DISULFID 411 422 BY SIMILARITY.
 FT DISULFID 416 431 BY SIMILARITY.
 FT DISULFID 433 442 BY SIMILARITY.
 FT DISULFID 450 461 BY SIMILARITY.
 FT DISULFID 455 467 BY SIMILARITY.
 FT DISULFID 469 478 BY SIMILARITY.
 FT CARBOHYD 244 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1295 AA; 144078 MW; 422AAD0A2DDEF3B4 CRC64;

Query Match 56.8%; Score 79; DB 1; Length 1295;
 Best Local Similarity 46.2%; Pred. No. 0.00024;
 Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 6 CXXKXGNGXCDXXCXXNNAACXXDXDC 31
 DB 542 CADQFANGVCNQECCNNECLYDGLDC 567

RESULT 15
 NTC4_MOUSE
 ID NTC4_MOUSE STANDARD; PRT; 1964 AA.
 AC P31695; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE [Contains: Transforming protein Int-3].
 GN NOTCH4 OR INT3 OR INT-3.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Callahan D., Callahan R.;
 RT "Mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells.";
 RL J. Virol. 66:2594-2599(1992).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294599; PubMed=9150355;
 RA Callahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4).";
 RL Oncogene 14:1883-1890(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene.";
 RL Development 122:2251-2259(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events.";
 RL J. Virol. 73:5166-5171(1999).
 RN [6]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 RP OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NI-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.


```

CC CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC CC proteolytical processing NICD is translocated to the nucleus.
CC CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
CC CC kidney, and at lower levels in the ovary and skeletal muscle. A
CC CC very low expression is seen in the brain, intestine, liver and
CC CC testis.
CC CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
CC CC embryonic development from 9.0 dpc.
CC CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC CC which is proteolytically cleaved by a furin-like convertase in the
CC CC trans-Golgi network before it reaches the plasma membrane to yield
CC CC an active, ligand-accessible form. Cleavage results in a C-
CC CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC CC ligand binding, it is cleaved by a TNF-alpha converting enzyme
CC CC (TACE) to yield a membrane-associated intermediate fragment called
CC CC notch extracellular truncation (NEXT). This fragment is then
CC CC cleaved by presenilin dependent gamma-secretase to release a
CC CC notch-derived peptide containing the intracellular domain (NICD)
CC CC from the membrane.
CC CC -1- PTM: Phosphorylated.
CC CC -1- DISEASE: Loss of the extracellular domain causes constitutive
CC CC activation of the Notch protein, which leads to hyperproliferation
CC CC of glandular epithelial tissues and development of mammary
CC CC carcinomas.
CC CC -1- SIMILARITY: Belongs to the NOTCH family.
CC CC -1- SIMILARITY: Contains 29 EGF-like domains.
CC CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC CC -1- SIMILARITY: Contains 5 ANK repeats.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; M80456; AAB38377.1; -.
DR DR EMBL; U43691; AAC52630.1; -.
DR DR EMBL; U43691; AAC52631.1; -.
DR DR EMBL; AF030001; AAB82004.1; -.
DR DR EMBL; AB016771; BAA32281.1; ALT_SEQ.
DR DR EMBL; AB016772; BAA32283.1; ALT_INIT.
DR DR EMBL; AB016773; BAA32284.1; ALT_INIT.
DR DR EMBL; AB016774; BAA32285.1; -.
DR DR PIR; A38072; TWMVT3.
DR DR PIR; T09059; T09059.
DR DR HSSP; P08709; 1BF9.
DR DR MGD; MGI:107471; Notch4.
DR DR InterPro; IPR002110; ANK.
DR DR InterPro; IPR000152; Asx_hydroxyl_S.
DR DR InterPro; IPR000742; EGF_2.
DR DR InterPro; IPR001881; EGF_Ca.
DR DR InterPro; IPR001438; EGF_II.
DR DR InterPro; IPR006209; EGF_like.
DR DR InterPro; IPR002049; laminin_EGF.
DR DR InterPro; IPR008297; Notch.
DR DR InterPro; IPR008800; Notch_dom.
DR DR Pfam; PF00023; ank; 6.
DR DR Pfam; PF00008; EGF; 27.
DR DR Pfam; PF00066; notch; 2.
DR DR PIRSF; PIRSF002279; Notch; 1.
DR DR PRINTS; PR00010; EGFBL00D.
DR DR PRINTS; PR00011; EGFLAMININ.
DR DR PRINTS; PR01452; NOTCH.
DR DR SMART; SM00248; ANK; 6.
DR DR SMART; SM00179; EGF_CA; 11.
DR DR SMART; SM00004; NL; 2.
DR DR PROSITE; PS0297; ANK_RBP_REGION; 1.
DR DR PROSITE; PS00088; ANK_REPEAT; 5.
DR DR PROSITE; PS00010; ASX_HYDROXYL; 11.
DR DR PROSITE; PS00022; EGF_1; 28.
DR DR PROSITE; PS01186; EGF_2; 21.

```

DR	PROSITE; PS50026; EGF_3; 27.
DR	PROSITE; PS01187; EGF_CA; 9.
KW	Receptor; Transcription regulation; Activator; Differentiation;
KW	Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW	Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
FT	SIGNAL 1 20 POTENTIAL.
FT	CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT	CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.
FT	CHAIN 1428 1964 NOTCH EXTRACELLULAR TRUNCATION.
FT	CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.
FT	DOMAIN 21 1443 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1444 1464 POTENTIAL.
FT	DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 21 60 EGF-LIKE 1.
FT	DOMAIN 61 112 EGF-LIKE 2.
FT	DOMAIN 115 152 EGF-LIKE 3.
FT	DOMAIN 153 189 EGF-LIKE 4.
FT	DOMAIN 191 229 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 231 271 EGF-LIKE 6.
FT	DOMAIN 273 309 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 311 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 352 388 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 389 427 EGF-LIKE 10.
FT	DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 623 656 EGF-LIKE 16.
FT	DOMAIN 658 686 EGF-LIKE 17.
FT	DOMAIN 688 724 EGF-LIKE 18.
FT	DOMAIN 726 762 EGF-LIKE 19.
FT	DOMAIN 764 800 EGF-LIKE 20.
FT	DOMAIN 803 839 EGF-LIKE 21.
FT	DOMAIN 841 877 EGF-LIKE 22.
FT	DOMAIN 878 924 EGF-LIKE 23.
FT	DOMAIN 926 962 EGF-LIKE 24.
FT	DOMAIN 964 1000 EGF-LIKE 25.
FT	DOMAIN 1002 1040 EGF-LIKE 26.
FT	DOMAIN 1042 1081 EGF-LIKE 27.
FT	DOMAIN 1083 1122 EGF-LIKE 28.
FT	DOMAIN 1126 1167 EGF-LIKE 29.
FT	REPEAT 1168 1208 LIN/NOTCH 1.
FT	REPEAT 1209 1242 LIN/NOTCH 2.
FT	REPEAT 1243 1282 LIN/NOTCH 3.

```

Query Match          56.8%; Score 79; DB 1; Length 1964;
Best Local Similarity 50.0%; Pred. No. 0.0036;
Matches 13; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY      6 CXXKXGNGXCDXXCNNNAACXXDXDC 31
      | | | | | | | | | | | | | |
Db      1255 CRDHFHNGHCXKGCNNNAECGWDGDC 1280

```

Search completed: May 19, 2004, 15:58:21
Job time : 0.785805 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 19, 2004, 15:54:24 ; Search time 2.98606 Seconds

(without alignments)
3275.577 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139

Sequence: 1 CDXXXXXXGNGXCDXXCNNAACXXDXDC 31

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	70.5	638	13	042372 brachydanio
2	97	69.8	585	5	Q9U0E2 tribolium c
3	93	66.9	1194	13	Q9W737 gallus gall
4	91	65.5	2531	5	016004 lytechinus
5	90	64.7	191	5	Q8T5Z0 caenorhabdi
6	90	64.7	191	5	Q8T5Z1 caenorhabdi
7	90	64.7	193	5	Q8T5G0 caenorhabdi
8	90	64.7	193	5	Q8T5Z2 caenorhabdi
9	90	64.7	194	5	Q8T5Z3 caenorhabdi
10	90	64.7	194	5	Q8T5Z3 caenorhabdi
11	90	64.7	1270	5	Q9GPM9 caenorhabdi
12	90	64.7	1441	5	Q9GPM9 caenorhabdi
13	89	64.0	2528	13	Q8AXP0 caenorhabdi
14	88	63.3	1476	13	Q90285 carassius a
15	87	62.6	2428	5	Q816X6 boophilus m
16	85	61.2	892	5	Q25243 lucilia cup

17	85	61.2	1455	5	Q86FJ9	Q86fj9 caenorhabdi
18	85	61.2	2516	11	Q7TQ52	Q7tq52 mus musculu
19	85	61.2	2526	11	Q7TQ51	Q7tq51 mus musculu
20	85	61.2	2531	11	Q8K428	Q8k428 mus musculu
21	85	61.2	2531	11	Q7TQ50	Q7tq50 mus musculu
22	85	61.2	2653	5	Q25253	Q25253 lucilia cup
23	83	59.7	762	13	Q42373	Q42373 brachydanio
24	83	59.7	2468	13	Q800E4	Q800e4 brachydanio
25	81	58.3	2524	5	Q9GPA5	Q9gpa5 branchiosto
26	79	56.8	1308	5	Q9GPM8	Q9gpm8 caenorhabdi
27	78	56.1	847	4	Q96N13	Q96n13 homo sapien
28	78	56.1	950	4	Q9TUL2	Q9tul2 homo sapien
29	78	56.1	3869	5	Q86PQ3	Q86pq3 cryptospori
30	77	55.4	260	13	Q9DEC9	Q9dec9 gallus gall
31	77	55.4	641	13	Q98U14	Q98u14 gallus gall
32	77	55.4	2352	5	Q61240	Q61240 halocynthia
33	76	54.7	752	13	Q42374	Q42374 brachydanio
34	74	53.2	40	5	Q9TXA3	Q9txa3 caenorhabdi
35	73	52.5	666	5	Q8SXI4	Q8sxi4 drosophila
36	73	52.5	666	5	Q9V553	Q9v553 drosophila
37	73	52.5	2447	13	Q13149	Q13149 fugu rubrip
38	71	51.1	123	5	Q8T5X3	Q8t5x3 caenorhabdi
39	71	51.1	123	5	Q8T5W3	Q8t5w3 caenorhabdi
40	71	51.1	124	5	Q8T5X0	Q8t5x0 caenorhabdi
41	71	51.1	124	5	Q8T5W6	Q8t5w6 caenorhabdi
42	71	51.1	124	5	Q8T5W5	Q8t5w5 caenorhabdi
43	71	51.1	126	5	Q8T5W4	Q8t5w4 caenorhabdi
44	71	51.1	127	5	Q8T5W9	Q8t5w9 caenorhabdi
45	71	51.1	127	5	Q8T5X2	Q8t5x2 caenorhabdi

ALIGNMENTS

RESULT 1
042372 ID 042372 PRELIMINARY; PRT; 638 AA.
AC 042372;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Notch receptor protein (Fragment).
GN NOTCH1B OR NOTCH1B.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Westin J., Lardelli M.;
RT "Three novel Notch genes in zebrafish: implications for vertebrate
RT Notch gene evolution and function.";
RL Dev. Genes Evol. 207:51-63(1997).
DR EMBL; Y10352; CAA71378.1; -.
DR HSSP; P00740; IEDM.
DR ZFIN; ZDB-GENE-990415-183; notch1b.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00022; EGF_1; 2.
KM ANK repeat; EGF-like domain; Receptor; Repeat.
FT NON_TER 1
SQ SEQUENCE 638 AA; 70800 MW; BCEA16D7561FA3D2 CRC64;

Query Match 70.5%; Score 98; DB 13; Length 638;
Best Local Similarity 51.6%; Pred. No. 1.1e-08;
Matches 16; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGXCDXXCNNAACXXDXDC 31
Db 86 CEIEQCYKXGNKICDSACNNYACDWDGDC 116

RESULT 2

Q9U0E2 PRELIMINARY; PRT; 585 AA.
AC Q9U0E2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Signal receptor protein (Fragment).
GN NOTCH.
OS Tribolium castaneum (Red flour beetle).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformata;
OC Tenebrionidae; Tribolium.
CX NCBI_TaxID=7070;
RN [1]

RP SEQUENCE FROM N.A.
RA Tautz D., Lardelli M., Westin J., Tamme R.;
RT "Embryonic expression of Tribolium Notch."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005083; CAB65469.1; -.

DR HSSP; P00740; 1EDM.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 12.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00179; EGF_CA; 5.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 6.
DR PROSITE; PS00022; EGF_1; 12.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS01187; EGF_CA; 4.
KW EGF-like domain; Receptor.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 585 AA; 63337 MW; 4CF7A5ID0820D048 CRC64;

Query Match 69.8%; Score 97; DB 5; Length 585;
Best Local Similarity 51.6%; Pred. No. 1.6e-08;
Matches 16; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGXCDXXCNNAACXXDXDC 31
Db 488 CNQKPCPVKRGNGRCDEGNTYACEFDGDC 518

RESULT 3

Q9W737 PRELIMINARY; PRT; 1194 AA.
ID Q9W737

AC Q9W737;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Notch-1 (Fragment).

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CX NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=93328644; PubMed=10402194;

RA Wakamatsu Y., Maynard T.M., Jones S.U., Weston J.A.;

RT "Numb localizes in the basal cortex of mitotic avian neuroepithelial

cells and modulates neuronal differentiation by binding to NOTCH-1.";

RL Neuron 23:71-81(1999).

DR EMBL; AF159231; AAD42893.1; -.

DR HSSP; P00740; 1EDM.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR001064; Crystallin.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00008; EGF; 1.

DR Pfam; PF00066; notch; 3.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 5.

DR SMART; SM00181; EGF; 1.

DR SMART; SM00004; NL; 3.

DR PROSITE; PS50088; ANK_REPEAT; 4.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.

DR PROSITE; PS00022; EGF_1; 2.

KW ANK repeat; Repeat.

FT NON_TER 1

SQ SEQUENCE 1194 AA; 130641 MW; 650380B8E6584974 CRC64;

Query Match

Best Local Similarity 66.9%; Score 93; DB 13; Length 1194;
Matches 15; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGXCDXXCNNAACXXDXDC 31
Db 92 CEIACASVAGNKCICGKCNHACGWDGDC 122

RESULT 4

ID 016004 PRELIMINARY; PRT; 2531 AA.
AC 016004;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Notch homolog.

OS Lytechinus variegatus (Sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;

OC Lytechinus.

CX NCBI_TaxID=7654;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97454256; PubMed=9310331;

RA Sherwood D.R., McClay D.R.;

RT "Identification and localization of a sea urchin Notch homologue:

insights into vegetal plate regionalization and Notch receptor

regulation.";

RL Development 124:3363-3374(1997).

DR EMBL; AF000634; AAB82088.1; -.

DR HSPF; P01132; 1EGF.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 21.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS01187; EGF_CA; 20.
DR PIRSF; PIRSF002279; Notch; 1.
DR ANK repeat; EGF-like domain; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match 65.5%; Score 91; DB 5; Length 2531;
Best Local Similarity 53.8%; Pred. No. 6.3e-07;
Matches 14; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 6 CXXXXXGXCDXXCNNAACXXDXDC 31
Db 1526 CLERYGNFCDEECNNIGCLYDGLDC 1551

RESULT 5
Q8T5Z0 PRELIMINARY; PRT; 191 AA.
AC Q8T5Z0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT847;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09703.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR00800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5FA7FB470 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 191;
Best Local Similarity 48.4%; Pred. No. 9e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXXXGXCDXXCNNAACXXDXDC 31
Db 117 CKKNCKALAGNGICDHCNYAECQFDGDC 147

RESULT 6
Q8T5Z1 PRELIMINARY; PRT; 191 AA.
AC Q8T5Z1;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB800;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491462; AAM09702.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR00800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20977 MW; 2EEB21B5FA46470 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 191;
Best Local Similarity 48.4%; Pred. No. 9e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXXXGXCDXXCNNAACXXDXDC 31
Db 117 CKKNCKALAGNGICDHCNYAECQFDGDC 147

RESULT 7
Q8STG0 PRELIMINARY; PRT; 193 AA.
AC Q8STG0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT Clp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxId=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK104, and HK105;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus *Caenorhabditis*.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491459; AAM09699.1; -.
DR EMBL; AF491461; AAM09701.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR006210; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PRO0011; EGFLAMININ.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 193 193
SQ SEQUENCE 193 AA; 21180 MW; 413C8AB647B5C540 CRC64;

```

Query Match          64.7%; Score 90; DB 5; Length 193;
Best Local Similarity 48.4%; Pred. No. 9.1e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Ddb      120  CKKNCALAGNGICDEDCNYAECQFDGGDC 150

QY      1  CDXXXCKXKGGXCDXXCENNACXXDXGDC 31
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      |||||

RESULT 8
Q8T5Z2  Q8T5Z2      PRELIMINARY;      PRT;      193 AA.
ID      Q8T5Z2;
AC      Q8T5Z2;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Glp-1 (Fragment).
OS      Caenorhabditis briggsae.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6238;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AF16;
RA      Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT      "Levels of DNA polymorphism vary with mating system in the nematode
RT      genus Caenorhabditis.";
RL      Genetics 0:0-0(2002).
DR      EMBL: AF491460; AAM09700.1; -.
DR      GO: GO:0016020; C:membrane; IEA.
DR      GO: GO:0005198; F:structural molecule activity; IEA.
DR      GO: GO:0030154; P:cell differentiation; IEA.
DR      InterPro: IPR000152; Asx_hydroxyl_S.
DR      InterPro: IPR006209; EGF_like.
DR      InterPro: IPR006210; IEGF.

```

```

DR InterPro; IPR002049; Laminin EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PRO0011; EGF_LAMININ.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 193 193
SQ SEQUENCE 193 AA; 21178 MW; 42135BB8E9BE02C5 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 193;
Best Local Similarity 48.4%; Pred. No. 9, 1e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 CDXXXCXXKKGXGXCXXXCNNAACXXDXDC 31
Db 119 CKKKCKALAGNGICDEDCCNYAECQFDGDC 149

RESULT 9			
ID	Q8T5Z3	PRELIMINARY;	PRT; 194 AA.
AC	Q8T5Z3;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, last annotation update)		
DE	Glip-1 (Fragment).		
OS	Caenorhabditis briggsae.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6238;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PB826;		
RA	Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;		
RT	"Levels of DNA polymorphism vary with mating system in the nematode		
RL	genus <i>Caenorhabditis</i> .";		
RL	Genetics 0:0-0(2002).		
DR	EMBL; AF491458; AAM09698.1; -.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0030154; P:cell differentiation; IEA.		
DR	InterPro; IPR000152; Asx_hydroxyl_S.		
DR	InterPro; IPR006209; EGF_like.		
DR	InterPro; IPR006210; IEGF.		
DR	InterPro; IPR002049; Laminin_EGF.		
DR	InterPro; IPR000800; Notch_dom.		
DR	Pfam; PF00008; EGF_3.		
DR	Pfam; PF00066; notch; 2.		
DR	PRINTS; PR00011; EGF_LAMININ.		
DR	PRINTS; PR01452; NOTCH.		
DR	SMART; SM00181; EGF_3.		
DR	SMART; SM00004; NL; 2.		
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.		
DR	PROSITE; PS00022; EGF_1; 3.		
DR	PROSITE; PS01186; EGF_2; 3.		
KW	EGF-like domain.		
FT	NON_TER 1		
FT	NON_TER 194 194		
SQ	SEQUENCE 194 AA; 21275 MW; 07313CBABD9C65C5 CRC64;		

```

Query Match          64.7%;  Score 90;  DB 5;  Length 194;
Best Local Similarity 48.4%;  Pred. No. 9.1e-08;
Matches 15;  Conservative 0;  Mismatches 16;  Indels 0;  Gaps 0;
QY      1 CDXXXXXXKXGNGXCDXXC NNAACXXDXGDC 31

```

Db 120 CKKNCCKALAGNGICDEDNCYAECCQFDGDC 150

RESULT 10

Q9GPM9 PRELIMINARY; PRT; 963 AA.

AC Q9GPM9; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor (Fragment).

GN GLP-1.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6238;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AF16;

RX MEDLINE=21100341; PubMed=11156985;

RA Rudel D., Kimble J.;

RT "Conservation of glp-1 regulation and function in nematodes.";

RL Genetics 157:639-654(2001).

DR EMBL; AF315555; AAG49317.1; -.

DR HSSP; P01132; 1EGF.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR006210; 1EGF.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00008; EGF; 4.

DR Pfam; PF00066; notch; 3.

DR PRINTS; PR00011; EGF_LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 4.

DR SMART; SM00181; EGF; 4.

DR SMART; SM00004; NL; 3.

DR PROSITE; PS50088; ANK_REPEAT; 3.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS00022; EGF_1; 4.

DR PROSITE; PS01186; EGF_2; 3.

KW ANK repeat; EGF-like domain; Receptor; Repeat; Transmembrane.

FT NON_TER

SQ SEQUENCE 963 AA; 106394 MW; 52BF80010E12FB57 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 963;

Best Local Similarity 48.4%; Pred. No. 3.9e-07;

Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCKKNGXCDXXCNNAACXXDXDC 31

Db 187 CKKNCCKALAGNGICDEDNCYAECCQFDGDC 217

RESULT 11

Q9GPN0 PRELIMINARY; PRT; 1270 AA.

AC Q9GPN0; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor.

GN GLP-1.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6238;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AF16;

RX MEDLINE=21100341; PubMed=11156985;

RA Rudel D., Kimble J.;

RT "Conservation of glp-1 regulation and function in nematodes.";

RL Genetics 157:639-654(2001).

DR EMBL; AF315554; AAG49316.1; -.

DR HSSP; P01132; 1EGF.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00008; EGF; 10.

DR Pfam; PF00066; notch; 3.

DR PRINTS; PR00011; EGF_LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 4.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00004; NL; 3.

DR PROSITE; PS50088; ANK_REPEAT; 3.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00010; ASX_HYDROXYL; 3.

DR PROSITE; PS00022; EGF_1; 10.

DR PROSITE; PS01186; EGF_2; 8.

DR PROSITE; PS01187; EGF_CA; 1.

KW ANK repeat; EGF-like domain; Receptor; Repeat; Transmembrane.

SQ SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 64.7%; Score 90; DB 5; Length 1270;

Best Local Similarity 48.4%; Pred. No. 5e-07;

Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCKKNGXCDXXCNNAACXXDXDC 31

Db 494 CKKNCCKALAGNGICDEDNCYAECCQFDGDC 524

RESULT 12

Q867Q2 PRELIMINARY; PRT; 1441 AA.

AC Q867Q2; 01-JUN-2003 (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor LIN-12.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6238;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22243075; PubMed=12356262;

RA Rudel D., Kimble J.;

RT "Evolution of discrete notch-like receptors from a distant gene duplication in Caenorhabditis.";

RL Evol. Dev. 4:319-333(2002).

RL EMBL; AF499438; AAP05763.1; -.

DR EMBL; AF499440; AAP05765.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

